

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 21:08:50 ; Search time 5576 Seconds
(without alignments)
10827.985 Million cell updates/sec

Title: US-10-099-704-1
Perfect score: 1393
Sequence: 1 ccgaagtgagcaatgag.....cgaaaaaaaaaaaaaaaa 1393

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1393	100.0	1393	6	BD236906	BD236906 Method fo
2	1393	100.0	1393	6	AR208625	AR208625 Sequence
3	214.8	15.4	2505	8	CPU312754	AJ312754 Claviceps
4	210.8	15.1	2182	8	CPU011963	AJ011963 Claviceps
5	210.8	15.1	2299	8	AY259840	AY259840 Claviceps
6	210.6	15.1	2694	8	CPU312753	AJ312753 Claviceps
7	205.2	14.7	1347	6	AR181512	AR181512 Sequence
8	205.2	14.7	1353	6	AR181513	AR181513 Sequence
9	194	13.9	2143	8	AY262013	AY262013 Balansa
10	194	13.9	2570	8	CPU312752	AJ312752 Claviceps
11	193.6	13.9	1783	8	AY262014	AY262014 Claviceps
12	182.6	13.1	1894	8	CLCDMAW	LJ9640 Claviceps f
13	176.6	12.7	1598	6	AR181515	AR181515 Sequence
14	176.6	12.7	6055	8	AY259839	AY259839 Neotyphod
15	176.6	12.7	6640	8	AY259837	AY259837 Epithloe
16	173.4	12.4	1908	6	AR181514	AR181514 Sequence
17	173.4	12.4	6702	8	AY259838	AY259838 Neotyphod
C 18	44.6	3.2	125020	9	AF429315	AF429315 Homo sapi
C 19	43.6	3.1	2000	6	AX655393	AX655393 Sequence
C 20	41.6	3.0	125020	9	AF429315	AF429315 Homo sapi
C 21	41	2.9	115368	9	AC135049	AC135049 Homo sapi
C 22	41	2.9	148054	9	AC026467	AC026467 Homo sapi
C 23	41	2.9	156899	2	AC103884	AC103884 Homo sapi
C 24	41	2.9	173980	9	AC116552	AC116552 Homo sapi
C 25	41	2.9	190590	2	AC135045	AC135045 Homo sapi
C 26	40.2	2.9	251271	2	AC129644	AC129644 Rattus no
C 27	40.2	2.9	253942	2	AC096037	AC096037 Rattus no
C 28	40.2	2.9	273754	2	AC120615	AC120615 Rattus no
C 29	38.8	2.8	2000	6	AX655393	AX655393 Sequence
C 30	38	2.7	295500	1	AP005954	AP005954 Bradyrhiz
C 31	37.8	2.7	2830	3	AF027972	AF027972 Nephila c
C 32	37.8	2.7	2830	6	AR091362	AR091362 Sequence
C 33	37.8	2.7	111738	2	AL592304	AL592304 Homo sapi
C 34	37.8	2.7	136116	2	AC110874	AC110874 Gallus ga
C 35	37.8	2.7	140207	9	AL356390	AL356390 Human DNA
C 36	37.8	2.7	166471	2	AC091052	AC091052 Homo sapi
C 37	37.8	2.7	202272	2	AC146292	AC146292 Gallus ga
C 38	37.8	2.7	239116	9	AP000426	AP000426 Homo sapi
C 39	37.6	2.7	112533	9	AL139097	AL139097 Human DNA
C 40	37.6	2.7	177521	2	AL161438	AL161438 Homo sapi
C 41	37.2	2.7	2765	8	AK064973	AK064973 Oryza sat
C 42	37.2	2.7	2997	8	AK122184	AK122184 Oryza sat
C 43	37.2	2.7	223951	2	EX005010	EX005010 Danio rer
C 44	36.6	2.6	1594	3	AY132363	AY132363 Spironucl
C 45	36.6	2.6	176099	9	AC108201	AC108201 Homo sapi

ALIGNMENTS

RESULT 1
BD236906
LOCUS BD236906 1393 bp DNA linear PAT 17-JUL-2003
DEFINITION Method for producing polypeptides in Aspergillus variant cells.
ACCESSION BD236906
VERSION BD236906.1 GI:33046676
KEYWORDS JP 2002533133-A/1.
SOURCE Aspergillus oryzae
ORGANISM Aspergillus oryzae
Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
REFERENCE 1 (bases 1 to 1393)
AUTHORS Christensen,B.E., Mcelligaard,H., Kaasgaard,S. and Lehmbeck,J.
TITLE Method for producing polypeptides in Aspergillus variant cells
JOURNAL Patent: JP 2002533133-A 1 08-OCT-2002;

NOVOZYMES AS
OS Aspergillus oryzae
PN JP 2002533133-A/1
PD 08-OCT-2002
PF 22-DEC-1999 JP 2000591212
PR 23-DEC-1998 DK PA 199801726, 27-MAY-1999 DK PA 199900745 PI
BOERN EGGERT CHRISTENSEN, HENRIK MOELGAARD, SVEND KAASGAARD, J.
JAN LEHMBECK
PC
C12N15/09, C12N1/15, C12N9/00, C12N9/20, C12N9/42, C12N15/01, C12P21/ PC
02//
PC (C12N1/15, C12P1/69), C12N15/00, C12N15/00
CC Method for producing polypeptides in Aspergillus variant cells
FH Key Location/Qualifiers
FT CDS (15)..(1328).
FEATURES
source
1. .1393
/organism="Aspergillus oryzae"
/mol_type="genomic DNA"
/db_xref="taxon:5062"
ORIGIN
Query Match 100.0%; Score 1393; DB 6; Length 1393;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCAACACTGCTGCCAAAGCCCTTCT 60
DB 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCAACACTGCTGCCAAAGCCCTTCT 60
QY 61 ACCTGTGTGAGTCAAGCCCTGAACTCTCGAAACAGGACCAACAAATGGTGGTATAGCA 120
DB 61 ACCTGTGTGAGTCAAGCCCTGAACTCTCGAAACAGGACCAACAAATGGTGGTATAGCA 120
QY 121 CAGCTCGGATGTTGGCACCATGATGGCGGGGCGCGGTATGATGTTCCAGCAGTACA 180
DB 121 CAGCTCGGATGTTGGCACCATGATGGCGGGGCGCGGTATGATGTTCCAGCAGTACA 180
QY 181 AGTTCTCTGATCCACCGTGAGTCAATCATCCCGCGGTGGTCCATACCAGAAAGG 240
DB 181 AGTTCTCTGATCCACCGTGAGTCAATCATCCCGCGGTGGTCCATACCAGAAAGG 240
QY 241 GTGAGCCCATGCACTGGAAGAGTCACTCAGCGTTTCGAGCTTTCCTTTGAGCTGAGCT 300
DB 241 GTGAGCCCATGCACTGGAAGAGTCACTCAGCGTTTCGAGCTTTCCTTTGAGCTGAGCT 300
QY 301 TCAATTACTCCCAATCACTACTACGTTTGCATTCGAGCCCTCGGTTCCCTGACGGGAA 360
DB 301 TCAATTACTCCCAATCACTACTACGTTTGCATTCGAGCCCTCGGTTCCCTGACGGGAA 360
QY 361 CGAAGGATGATCCATTCAACACCCAGGCAATCAGGCTGTTCTCCAGGACCTCAAGGCA 420
DB 361 CGAAGGATGATCCATTCAACACCCAGGCAATCAGGCTGTTCTCCAGGACCTCAAGGCA 420
QY 421 TGGTTCAGGGCTTGACCTGGAGTGTTCGATCAATTCATTAAGCAATTTGGTTCGG 480
DB 421 TGGTTCAGGGCTTGACCTGGAGTGTTCGATCAATTCATTAAGCAATTTGGTTCGG 480
QY 481 AGGAAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGTCTTCAAGACACAGA 540
DB 481 AGGAAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGTCTTCAAGACACAGA 540
QY 541 ACAAACTGGCAGCGGATCTGGAGCCATCTGGGATATTCTTTGAAGACCTACATCTACC 600
DB 541 ACAAACTGGCAGCGGATCTGGAGCCATCTGGGATATTCTTTGAAGACCTACATCTACC 600
QY 601 CGCGGATCAAGTGCATCGGACCGGACCCCAAGAGAGAGTCAATGTTTGAAGCAATCA 660
DB 601 CGCGGATCAAGTGCATCGGACCGGACCCCAAGAGAGAGTCAATGTTTGAAGCAATCA 660
QY 661 AGGCTCCGACAAAGTTGGCAAGTTGCCACTCCACTGGCAATCTCCAGGAGGTTTATAG 720
DB 661 AGGCTCCGACAAAGTTGGCAAGTTGGCAAGTTGCCACTCCACTGGCAATCTCCAGGAGGTTTATAG 720

QY 721 CTGAGCGAGCACCACCCCTCTCTCGGCACATTTCTCTCATCGATTTGGTCAAGCCGTCG 780
DB 721 CTGAGCGAGCACCACCCCTCTCTCGGCACATTTCTCTCATCGATTTGGTCAAGCCGTCG 780
QY 781 AGTCCCGAATCAAGGTCTACTGTATGAAAGCGCAGCTCGACCTGGCTCCATCGAAGTA 840
DB 781 AGTCCCGAATCAAGGTCTACTGTATGAAAGCGCAGCTCGACCTGGCTCCATCGAAGTA 840
QY 841 TTTGGACTCTCAACCGGCGAGCGGAACGATCCAGAGACACTGGATGGTCTCGATCGCGTGA 900
DB 841 TTTGGACTCTCAACCGGCGAGCGGAACGATCCAGAGACACTGGATGGTCTCGATCGCGTGA 900
QY 901 GGGAGCTGTGGCAGCTATTGCCCGTCAAGGAGGCTGTGTCCACTGCCGAACTGCTTT 960
DB 901 GGGAGCTGTGGCAGCTATTGCCCGTCAAGGAGGCTGTGTCCACTGCCGAACTGCTTT 960
QY 961 ACAGCGGGGTACTCCTCACCGCAGGAGCAGTCCCCTTCATTATAAATTTTACCTTGTCTC 1020
DB 961 ACAGCGGGGTACTCCTCACCGCAGGAGCAGTCCCCTTCATTATAAATTTTACCTTGTCTC 1020
QY 1021 CTAAGAGCGCACTTCCCGAACCAAGATCTATTTCCCTGCTTTTGGGCGAAGACGACAAA 1080
DB 1021 CTAAGAGCGCACTTCCCGAACCAAGATCTATTTCCCTGCTTTTGGGCGAAGACGACAAA 1080
QY 1081 CCATCGCGGAGGATTTGGCCACCTCTTTGAGAGCAGAGGTTGGGTGGCTTGGCTAAGA 1140
DB 1081 CCATCGCGGAGGATTTGGCCACCTCTTTGAGAGCAGAGGTTGGGTGGCTTGGCTAAGA 1140
QY 1141 GCTATCCAGCGGATTTGGCATCTCTACTATCCCGATGTCGACCTCCAGACCGCAATCACC 1200
DB 1141 GCTATCCAGCGGATTTGGCATCTCTACTATCCCGATGTCGACCTCCAGACCGCAATCACC 1200
QY 1201 TGGAGGCTGGATCTCTCTCTTACAGGGGAAACCGTACATGAGTGTGACTCTCC 1260
DB 1201 TGGAGGCTGGATCTCTCTCTTACAGGGGAAACCGTACATGAGTGTGACTCTCC 1260
QY 1261 ATACCTTCGAAGCGTTCAGTCTGCTGCCAGGAGGTTGGCTATGTCAGATGGCCACA 1320
DB 1261 ATACCTTCGAAGCGTTCAGTCTGCTGCCAGGAGGTTGGCTATGTCAGATGGCCACA 1320
QY 1321 ATCTTAGGACTAGTTTATCCCTTCACTATCCATCCCTGTCATGTCGATGGTTCGAAAAA 1380
DB 1321 ATCTTAGGACTAGTTTATCCCTTCACTATCCCTGTCATGTCGATGGTTCGAAAAA 1380
QY 1381 AAAAAAAAAAAAAA 1393
DB 1381 AAAAAAAAAAAAAA 1393
RESULT 2
LOCUS AR208625 1393 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383781.
ACCESSION AR208625
VERSION AR208625.1 GI:21509828
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1393)
AUTHORS Christensen, B. Eggert., Mollgaard, H., Kaasgaard, S. and Lehmbek, J.
TITLE Methods for producing polypeptides in aspergillus mutant cells
JOURNAL Patent: US 6383781-A 1 07-MAY-2002;
FEATURES Location/Qualifiers
source
1. .1393
/organism="unknown"
/mol_type="unassigned DNA"

Query Match 100.0%; Score 1393; DB 6; Length 1393;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CCGAAAGCTGAGCAATGGAGATCTCCAAGAAAGCAGCAACACTGCTGCCAAAGCCCTTCT	60
DB	1	CCGAAAGCTGAGCAATGGAGATCTCCAAGAAAGCAGCAACACTGCTGCCAAAGCCCTTCT	60
QY	61	ACGTGCTGAGTCAAGCCCTGAACTCTCGAAGCAAGGACCAACAAAATGGTGGTAGCA	120
DB	61	ACGTGCTGAGTCAAGCCCTGAACTCTCGAAGCAAGGACCAACAAAATGGTGGTAGCA	120
QY	121	CAGCTCCGATGTTTGGCACCATGATCGGGGGCCGGCTATGATGTTTACGCACAGTACA	180
DB	121	CAGCTCCGATGTTTGGCACCATGATCGGGGGCCGGCTATGATGTTTACGCACAGTACA	180
QY	181	AGTTCTCTGTATCCACCGTGAGGTGATCATTCGGCGGTTGGTTCATACCCAGAAAAGG	240
DB	181	AGTTCTCTGTATCCACCGTGAGGTGATCATTCGGCGGTTGGTTCATACCCAGAAAAGG	240
QY	241	GTCAAGCCATGCACTGGAAGAGTCATCTCACAGCTTTCGACATCTCTTTCGAGCTGAGCT	300
DB	241	GTCAAGCCATGCACTGGAAGAGTCATCTCACAGCTTTCGACATCTCTTTCGAGCTGAGCT	300
QY	301	TCAATTACTCCAAATCACTACTACGTTTGCAATTCGAGCCCTTCGTTCCCTGACGGGAA	360
DB	301	TCAATTACTCCAAATCACTACTACGTTTGCAATTCGAGCCCTTCGTTCCCTGACGGGAA	360
QY	361	CGAAGGATGATCATTCACACCCAGGCAATCAGGCTGTCTTCAGGACCTCAAGGCCCA	420
DB	361	CGAAGGATGATCATTCACACCCAGGCAATCAGGCTGTCTTCAGGACCTCAAGGCCCA	420
QY	421	TGGTTTCAGGGCTTGACCTCGAATGGTTTCGATCATTCACATAAAGCATTTGGTTCGTTCCG	480
DB	421	TGGTTTCAGGGCTTGACCTCGAATGGTTTCGATCATTCACATAAAGCATTTGGTTCGTTCCG	480
QY	481	AGGAAGGGCTCGGACTCTCTAGTTCGAGATTTGAGATCCCGCTTCCTCAAGACACAGA	540
DB	481	AGGAAGGGCTCGGACTCTCTAGTTCGAGATTTGAGATCCCGCTTCCTCAAGACACAGA	540
QY	541	ACAAATCGGAGCCGATCTCGAGCCATCTCGCGCATTTGTCTTGAAGACCTACATCTACC	600
DB	541	ACAAATCGGAGCCGATCTCGAGCCATCTCGCGCATTTGTCTTGAAGACCTACATCTACC	600
QY	601	CGGGATCAAGTCGATCGGACCGGGACCCCAAAGAGAGACTCATGTTTGACGCAATCA	660
DB	601	CGGGATCAAGTCGATCGGACCGGGACCCCAAAGAGAGACTCATGTTTGACGCAATCA	660
QY	661	AGGCTCCGACAAAGTTTGGCAAAAGTTTCCCACTCCACTGGCAATCTCTCGAGGAGTTATAG	720
DB	661	AGGCTCCGACAAAGTTTGGCAAAAGTTTCCCACTCCACTGGCAATCTCTCGAGGAGTTATAG	720
QY	721	CTGAGCAGACCCACCTCTCTGGCACTTTCTCTCATCGGATTTGGTCAAGCCGTCG	780
DB	721	CTGAGCAGACCCACCTCTCTGGCACTTTCTCTCATCGGATTTGGTCAAGCCGTCG	780
QY	781	AGTCCCGAAATCAAGGTCTACTGTATGGAACGCGAGCTCGACCTCGGCTCCATCGAAGGTA	840
DB	781	AGTCCCGAAATCAAGGTCTACTGTATGGAACGCGAGCTCGACCTCGGCTCCATCGAAGGTA	840
QY	841	TTTGGACTCTCAACGGCGGACGGAAACCATCCAGAGACACTGGATGGTTCGATCGCGTGA	900
DB	841	TTTGGACTCTCAACGGCGGACGGAAACCATCCAGAGACACTGGATGGTTCGATCGCGTGA	900
QY	901	GGGAGCTGTGGCAGCTATTGCCCTGACGAGGGTCTGTGTCCACTGCCGAACCTGCTTTT	960
DB	901	GGGAGCTGTGGCAGCTATTGCCCTGACGAGGGTCTGTGTCCACTGCCGAACCTGCTTTT	960
QY	961	ACGAGCCGGGTACTCTACCCGAGGAGCAGCTCCCTTCAATTATAATTTTACCTGTCTC	1020
DB	961	ACGAGCCGGGTACTCTACCCGAGGAGCAGCTCCCTTCAATTATAATTTTACCTGTCTC	1020
QY	1021	CTAAAAGCGACTTCCGAAACCAAGATCTATTTCCTGTCTTTTGGCAGAACGACAAA	1080
DB	1021	CTAAAAGCGACTTCCGAAACCAAGATCTATTTCCTGTCTTTTGGCAGAACGACAAA	1080

Qy	1081	CCATCGCGGAAGGATTTGGCCACCTTCTTTGAGAGCAGAGGTTGGGCTGGCTTGGCTAAGA	1141
Db	1081	CCATCGCGGAAGGATTTGGCCACCTTCTTTGAGAGCAGAGGTTGGGCTGGCTTGGCTAAGA	1140
Qy	1141	GCTATCCAGCGGATTTGGCATCTACTATCCCGATGTGACCTGCAGACCGCAAATCACC	1200
Db	1141	GCTATCCAGCGGATTTGGCATCTACTATCCCGATGTGACCTGCAGACCGCAAATCACC	1200
Qy	1201	TGCAGGCGTGGATCTCTCTTTCTTTACAAGGGGAAAAAACCGTTCATCATGAGTGTGTACCTCC	1260
Db	1201	TGCAGGCGTGGATCTCTCTTTCTTTACAAGGGGAAAAAACCGTTCATCATGAGTGTGTACCTCC	1260
Qy	1261	ATACCTTCGAAGCGTTTCAGTGTCTGCTGCCAGAGGTTGGCTATGTGTTCAGATGGCCACA	1320
Db	1261	ATACCTTCGAAGCGTTTCAGTGTCTGCTGCCAGAGGTTGGCTATGTGTTCAGATGGCCACA	1320
Qy	1321	ATCCTTAGGACTAGTTTATCCCTTCATTCATGCATCCGTTGAATGTTGGTGCAGAAAA	1380
Db	1321	ATCCTTAGGACTAGTTTATCCCTTCATTCATGCATCCGTTGAATGTTGGTGCAGAAAA	1380
Qy	1381	AAAAAAAAAAAAA	1393
Db	1381	AAAAAAAAAAAAA	1393
RESULT 3			
LOCUS	CPU312754	2505 bp	DNA linear PLN 24-APR-2001
DEFINITION	Claviceps purpurea partial cpd1 gene for putative dimethyl-allyl-tryptophan-synthase, exons 1-3, strain T5.		
ACCESSION	AJ312754		
VERSION	AJ312754.1	GI:13810199	
KEYWORDS	cpd1 gene; dimethyl-allyl-tryptophan-synthase.		
SOURCE	Claviceps purpurea (ergot fungus)		
ORGANISM	Claviceps purpurea		
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.		
AUTHORS	Correia, T. and Tudzynski, P.		
TITLE	Molecular analysis of dimethyl-allyl-tryptophan-synthase-genes		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2505)		
AUTHORS	Tudzynski, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-APR-2001) Tudzynski P., Institut fuer Botanik, Westfaelische Wilhelms-Universitaet, Schlossgarten 3, 48149 Muenster, GERMANY		
FEATURES	Location/Qualifiers		
source	1..2505		
	/organism="Claviceps purpurea"		
	/mol_type="genomic DNA"		
	/strain="T5"		
	/db_xref="taxon:5111"		
	/country="Germany"		
	1064..2505		
gene	/gene="cpd1"		
	join(1064..2229,2295..2351,2406..>2505)		
CDS	/gene="cpd1"		
	/functions="isoprenylation of tryptophane"		
	/codon_start=1		
	/product="putative dimethyl-allyl-tryptophan-synthase"		
	/protein_id="CAC37397.1"		
	/db_xref="GI:13810200"		
	/db_xref="SPTREMBL:Q9C140"		
	/translation="MSTAKDFGNGVYBILSLIFDPSPNEQRLMWHSTAPMPFAAMLDNNA		
	GYNIHDQYRHILGIFKKHIIIFPLGVPTQDKERWLISITRCGLPLELSINCTDSVYRVL		
	YEPINVEITKQKPNFLTAIWASVQKLAQICAGIDLEWFSYFKDELTLDBSEATLQSL		
	NELVKEQITKCNALDLKDSQALKYVFPHLKSIATGSKTDLIIDSVPFLKSQKH		
	SIQPAQVLCIDYVSRNRHSAESCHIALHARLISCDLIDPAKSRVXIVLLEKTVLSV		
	MYTLUGQRQRDASTWMDLMDRELISLKVPTGHLEYPKGYLELGEIENEQQLPSVA		
	NETLHHNNPPEPOVYFTFGMDNAEISNALTIPFORHGFDMMAKYRVFLQDSYFTH		
	DFESLNTLHYIISFSPVWPEQSTELYQHGLRKLNFLVADSPISFDAYR"		
exon	1064..2229		

```

/ gene="cpd1"
/ number=1
2230.2294
/ gene="cpd1"
/ number=1
2295.2351
/ gene="cpd1"
/ number=2
2352.2405
/ gene="cpd1"
/ number=2
2406.>2505
/ gene="cpd1"
/ number=3

ORIGIN
Query Match 15.4%; Score 214.8; DB 8; Length 2505;
Best Local Similarity 52.3%; Pred. No. 1.7e-52;
Matches 586; Conservative 2; Mismatches 499; Indels 33; Gaps 4;
Qy 57 TTCTACGTGCTGAGTCAAGCCCTGAACTCTCGAAGACGACACACAAAATGGTGGTAT 116
Db 1097 TACGAAATTTTGAGTCTGATTTTGGATTTCCTCCAGCAATGAACAGCGACTATGGTGGCAC 1156
Qy 117 AGCAGACCTCGATGTTTGCACCATGATGCGGGGCGGCTATGATGTTCAACGCACAG 176
Db 1157 AGTACCGCGCTATGTTTGGCGGATGCTTGACAATGCCGCTACATATCCAGACCAA 1216
Qy 177 TACAAGTTCCTCTGTATCCACCGTGAAGTCAATCCCGCGGTGGTGGTCCATACCCAGAA 236
Db 1217 TACCGGCATCTGGGCAATTTTCAAGAAGCACATATCCCTTTTCTTGGTGTCTATCCAACA 1276
Qy 237 AAGGTCAGCCCATGCACTCGAAGATCATCTCACAGCTTCGAGTCTCTTTCGAGCTG 296
Db 1277 AAAGACAGGAAA---GATGGCTCAGATCTCCACAGATGGCCCTTCTTTGGAAATG 1333
Qy 297 AGCTTCAATTAATCCAAATCACTACTACGGTTCGATTCGAGCCCTCGGTTCCCTGAGC 356
Db 1334 AGCTTGAATTTGACCGACTCTGTTTTCGATATACATACGAGCCATCAATGAAGTACG 1393
Qy 357 GGAACGAGGATGATCATTCACACCCAGGCAATCAGGCTGTTCTCAGCAGCTCAAG 416
Db 1394 GGGACGAGAAAGATCGTTCAATACATTTGCAATTTATGGCAAGTGTCCAAAACCTGCA 1453
Qy 417 GCATGTTCCAGGGTTCAGCTGGAATGGTTCGATCATTTCACTAAAGCATTTGGTCGTT 476
Db 1454 CAGATTCAAGCGGTATCGACTTGGAGTGGTTAGTTACTTCAAAGATGAGTGAAGCTG 1513
Qy 477 TCGAGGAGAGAGCTCGGACTCTGTAGATCGAGATATTGAGATCCCTCTTCAAGACA 536
Db 1514 GACGAATCGGAATCGGCCACACTTCAAAGTAATGAGCTGGTCAAGGAGCAGATAAGAGC 1573
Qy 537 CAGAACAACTCGCAGCCGATCTGGAGCCATCTGGGATATTGCTTTGAAGACCTACATC 596
Db 1574 CAGAACAACTTGGCTTGGATCTCAA---AGAAAGCCAGTTTCGGCTCAAAGTTTACTTC 1630
Qy 597 TACCGCGGATCAAGTCGATCGGACCGGAGCCCAAGGAGAGACTCATGTTTGAACGC- 655
Db 1631 TATCCGCATCTTAAATCGATCGGACCGGCAAAATCCACAGCATCTCATCTTCGACTCT 1690
Qy 656 -----AATCAAGCTCGCAGACAGTTTGGCAAGTTTGCACCTCCACCTGCAATCTCCGAG 710
Db 1691 GTGTTCAACTGTGCGAGAGAGATGACAGCATACAGCCCGCTTCAGGTATTGTCGCAC 1750
Qy 711 GAGTTT-----ATAGTTCAGCGAGCAGCCCAACCTCTCTCGGCCAC 749
Db 1751 TATGTTTTCGGCGGAAATCATTCGCGAGAATCAGACCAACATAGACCTCATACGCGCA 1810
Qy 750 TTTCTCTCATGCGATTTGGTCAAGCCGTCGAGTCCCGAATCAAGTCTACTGTATGGA 809
Db 1811 CTCTTGTATGCGATTTGATCGATCCCGCAAGTCTCGGCTCAAGATATACCTGCTGAG 1870
Qy 810 CGCCAGCTCGACCTGCGCTCCATCGAAGGATTTTGGACTCTCAACGGCGCAGCAACGAT 869

```

```

Db 1871 AAGACGGTCTCGTTGTCGGTATGGAAGATCTGTGGACGCTGGGCGGCCAACGAGTCGAC 1930
Qy 870 CCAGAGACACTGGATGGTCTGGATCGCTGAGGAGCTGTGGCAGCTATTGCCCGTCAAG 929
Db 1931 GCATCCACCATGAGCGGCTTGACATGCTTCGCGAGCTCTGGAGCTTCTAAAGATTCCC 1990
Qy 930 GAGGTCTGTCTCCACTGCCGAACTGCTTTTACAGCCGGGTACTCTCACGCGAGGAGAG 989
Db 1991 ACTGCCACATTCGAGTATCCAAAAGGGTATCTGGAATTGGGAGAAATCCGACGAGCAG 2050
Qy 990 CTCCTCTCATTAATAATTTTACCTTCTCTAAAGCGCACCTTCCTCCGACACACATC 1049
Db 2051 CTCCTATCCATGGCCAACTATACACTGCACCAACAAACCCCATGCTGAGCCCGAGTG 2110
Qy 1050 TATTTCCCTGCTTTTGGGAGAACGACAAACCAATCCGGAAGGATTGGCCCACTTTT 1109
Db 2111 TATTTACGGTTCCTGGCATGATGACCGCGMAATCAGCAATGCTTTGACCATCTTCTTC 2170
Qy 1110 GAGGACAGAGTTCGGGTGGCTTGGCTAAGAGCTATCCAG 1149
Db 2171 CAACGTACCGATTTCAGACATGCGGAAAAGTACCGAG 2210

RESULT 4
CPU011963 Claviceps purpurea dl gene. 2182 bp DNA linear PLN 14-FEB-2003
LOCUS Claviceps purpurea dl gene.
DEFINITION Claviceps purpurea dl gene.
ACCESSION AJ011963
VERSION AJ011963.1 GI:4499836
KEYWORDS D1 gene; dimethyl-allyl-tryptophan-synthase.
SOURCE Claviceps purpurea (ergot fungus)
ORGANISM Claviceps purpurea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.

REFERENCE
1 Tudzynski, P., Holter, K., Correia, T., Arntz, C., Grammel, N. and
Keller, U.
Evidence for an ergot alkaloid gene cluster in Claviceps purpurea
Mol. Gen. Genet. 261 (1), 133-141 (1999)
99168777
PUBMED 10071219
Tudzynski, P.
Direct Submission
AUTHORS Tudzynski, P.
TITLE Submitted (13-OCT-1998) Tudzynski P., Westfäelische
JOURNAL Wilhelms-Universität, Institut fuer Botanik, Schlossgarten 3,
Muenster, D-48149, GERMANY
REFERENCE 2 (bases 1 to 2182)
Tudzynski, P.
Direct Submission
AUTHORS Tudzynski, P.
TITLE Submitted (13-OCT-1998) Tudzynski P., Westfäelische
JOURNAL Wilhelms-Universität, Institut fuer Botanik, Schlossgarten 3,
Muenster, D-48149, GERMANY

FEATURES
source
1. .2182
/organism="Claviceps purpurea"
/mol_type="genomic DNA"
/strain="P1"
/db_xref="taxon:5111"
512..1977
/gene="dl"
/join("512..1677,1743..1864,1919..>1977)
/join("512..1677,1743..1864,1919..1977)
/codon_start=1
/product="dimethyl-allyl-tryptophan-synthase"
/protein_id="CA39314.1"
/db_xref="GI:4499837"
/db_xref="SPTREMBL:O94204"
/translation="MSTAKDPNGVYEIILSFDFPSBQRLWWHSTAMPFAMLDNA
GYNIHQYRLGIFKKHIIPLGVPTKDKRWLSILTCGLPLELSLNTSSVRYT
YEPINVTGKEDTNTLAIMTSVKLOAIQAGIDLEWFSYFKDELTLDESSTATQ
NELVQIOLKONKALDLKESOPALAKVFEHLKSIATGKSTHDLIFDSVLKLSKH
SIQAPQVCLDYVSRNHNSEVDQGHARLLSCDLIDPAKSRVKIYLLKXTVLSV
MEDTLGGORVDASTMDGLMLRELWLSLKVPTGHLKPTGKLELGEIIPNEQLPSNA
NYTLHNDPMPPEQVYTFVGMNDABEISNALTIFQFHGFDDMAKNRYFLQDSYPH
DFESLNYLHAYISFSYRRNKPYLSVYLHTFETGRWFEVADSPISFDAYRRCELSTK"

```

exon	<512..1677	
	/gene="di"	
	/number=1	
intron	1678..1742	
	/gene="di"	
	/number=1	
exon	1743..1864	
	/gene="di"	
	/number=2	
intron	1865..1918	
	/gene="di"	
	/number=2	
exon	1919..>1977	
	/gene="di"	
	/number=3	
ORIGIN		
Query Match 15.1%; Score 210.8; DB 8; Length 2182;		
Best Local Similarity 52.2%; Pred. No. 2.7e-51;		
Matches 585; Conservative 0; Mismatches 502; Indels 33; Gaps 4;		
QY	57	TTCTACGTCGTGAGTCAAGCCCTGACCTCTCGAACAGGACACACAAAATGGTGTAT 116
DB	545	TACGAAATTTGAGTCTGATCTTTGATTTCCCGACCAATGAACAGGACTATGGTGGAC 604
QY	117	AGCACAGCTCCGATGTTTGCACCATGATGGGGGGGGCGGCTATGATGTTCCACGCACAG 176
DB	605	AGTACCGCGCTATGTTTGGCGGATGCTTGACAATGCCGCTACAAATATCCACGACCAA 664
QY	177	TACAACTTCTCTGTATCCACCGTGAAGTATCATCCCGGGGCTGGTGGTCCATACCCAGAA 236
DB	665	TATCGGCATCTGGGCATTTTCAAGAAAGCAATATCCCTTTTCTTGGTGTCTATCCAACA 724
QY	237	AAGGGTCAGCCCATGCACTGGAAGAGTCTCTCACAGCTTCGGACTTCCTTTTCGAGGTG 296
DB	725	AAGAACAAGMAA---GATGGCTCAGATCCTCACCAGATGCGCTTCCTTTGGAATG 781
QY	297	AGCTTCAATCTCAAAATCACTACTACGTTTGGATTCGAGCCCTCGGTTCCCTGACG 356
DB	782	AGCTTCAATTTGTAACCGACTCTGTCTGATATACATACGAGCCCATCAATCAAGTGAAG 841
QY	357	GGAAAGAGGATGATCCATTCACACCCAGGCAATCAGGCTGTCTCCAGACCTCAAG 416
DB	842	GGACGAGGAAGATACGTTCAATACATTTGGCAATTTATGACAAGTGTCCAAAATCTGCA 901
QY	417	GCATGGTTCCAGGGCTTGACCTCGAATGTTTCGATCATTTCACTAAAGCATTTGGTCGTT 476
DB	902	CAGATTCGAAGCGGTATCGACTTTGGAGTGTGTTAGTTACTTCAAAGATGAGCTGACGCTG 961
QY	477	TCGAGGAGAGAGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCCGTCTTCAGACA 536
DB	962	GACGAATCGGAATCGGCCACACTTCAAAGTAATGAGCTGTCAAGGAGCAGATAAAGAGC 1021
QY	537	CAGAAACAACTGGCAGCCGATCTGGAGCCATCTGGGGATATTGTTCTTGAAGACCTACATC 596
DB	1022	CAGAACAGTTAGCCTTGGATCTCAA---AGAAGCCAGTTTCGGCTCAAAGTTTACTTTC 1078
QY	597	TACCGCGGATCAAGTTCGATTCGGAACGGGACCCGGAACCCAAAAGAGAGACTCATGTTGA ---C 653
DB	1079	TATCCGCATCTTAAATCGATCGGACCCGGCAATCCACACGATCTGATCTTTCGACTCC 1138
QY	654	GCAATCAAGCTCCCGACAGTTTGGCAAGTTGGCAAGTTGCCACTCCACTGGCAATCTTCGAGGAG 713
DB	1139	GTGTTGAAGCTGTCCGAGAAGCATGACGATACAGCCCGCGTTCCAGGTATTTGTCGAC 1198
QY	714	TTTTAT-----AGCTGAGCGAGCACCCACCCCTCTCTCGGCCAC 749
DB	1199	TATGTTTCGGCGGAAATCATTCGCGAGAAGTAGACCAACAGGAGCCCTACATCGGCA 1258
QY	750	TTTCTCTCATGCGATTTGTCAGCCGTCGAGTCCGATCCGATCCGATCTACTGTATGGAA 809
DB	1259	CTCTTGTGTCGATTTGATTCGATCCCGCCAAAGTCTCGCGTCAAGATATACCTACTGGAG 1318

QY	810	CGCCAGCTCGACCTGGCCTCCATCGAAGGTATTTGGACTCTCAACGGGCGACGACGAT 869
DB	1319	AAGACGGTCTCGTTGTCGGTGTGGAAGATCTGTGGACGCTGGCGGCGCAACGATCGAC 1378
QY	870	CCAGAGACACTGGATGCTCTGGATCGCTGAGGAGCTGTGGCAGCTATTGCCCGTCAAG 929
DB	1379	GCATCCACCATGGACGGCTTGACATGCTTCGGAGCTCTGGAGCTGTCTGAAAGTTCCC 1438
QY	930	GAGGTCGTGTCCACTGCCGAACTGCTTTTACGNGCCGGGTACCTCACCGCAGGACAG 989
DB	1439	ACTGGCCACTTGGAGTATCCAAAAGGTATCTGGAATTGGAGAAATTCGACAGCAG 1498
QY	990	CTCCCTTCAATATAAATTTTACCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1049
DB	1499	CTCCCATCCATGGCCAACTATACATCGACACACAAACGACCCATGCTGAGCCCGAGTG 1558
QY	1050	TATTTCCCTGCTTTTGGGAGACGACAAAACCATCTCGGGAAGATTGGCCACTTTT 1109
DB	1559	TATTTACGGTTTTCGGCATGAATGACGCCGAAATCAGCAATGCTTTGACCATCTTCTTC 1618
QY	1110	GAGAGCAGAGTTCGGGTGGCTTGGCTAAGAGCTATCCAG 1149
DB	1619	CAAGCTCAGGATTGAGGACATGGCGAAAATCTACCGAG 1658
RESULT 5		
AY259840		
LOCUS		
DEFINITION		
Claviceps purpurea dimethylallyltryptophan synthase (dmaW) gene,		
dmaW-1 allele, complete cds, alternatively translated.		
ACCESSION		
AY259840		
VERSION		
AY259840.1 GI:32402653		
KEYWORDS		
SOURCE		
ORGANISM		
Claviceps purpurea (ergot fungus)		
Claviceps purpurea		
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;		
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.		
REFERENCE		
AUTHORS		
1 (bases 1 to 2299)		
TITLE		
The determinant step in ergot alkaloid biosynthesis by a grass		
endophyte		
JOURNAL		
Fungal Genet. Biol. (2004) In press		
REFERENCE		
2 (bases 1 to 2299)		
AUTHORS		
Wang, J. and Schardl, C.		
TITLE		
Direct Submission		
JOURNAL		
Submitted (17-MAR-2003) Plant Pathology, University of Kentucky,		
201 Plant Sciences Bldg., Lexington, KY 40546-0312, USA		
FEATURES		
Location/Qualifiers		
1..2299		
/organism="Claviceps purpurea"		
/mol_type="genomic DNA"		
/isolate="ATCC20102"		
/specific_host="Secale cereale"		
/db_xref="ATCC:20102"		
/db_xref="taxon:5111"		
<600..>2116		
/gene="dmaW"		
gene		
allele="1"		
join(<600..1816,1882..2003,2058..>2116)		
/gene="dmaW"		
allele="1"		
/product="dimethylallyltryptophan synthase"		
join(600..1816,1882..2003,2058..2116)		
/gene="dmaW"		
allele="1"		
/EC_number="2.5.1.34"		
note="aromatic prenyltransferase; L-tryptophan		
dimethylallyl transferase;		
4-(gamma,gamma)-dimethylallyltryptophan synthase; Dmat		
synthase; 4-(3-methylbut-2-enyl)-L-tryptophan synthase;		
dimethylallyl-pyrophosphate; most likely true start codon		
might be the short isoform based on comparison with		
GenBank Accession number U39640; both long and short		

isoforms give rise to functional protein as confirmed by heterologous expression in yeast"

/codon_start=1

/evidence=experimental

/product="dimethylallyltryptophan synthase long isoform"

/protein_id="AAP81209.1"

/db_xref="GI:32402654"

/translation="MCTETDVSUSCPKVAAMSTAKDPNGVVEILSLIFDPPSNEORLWHSSTAPFAAMLDNAGYNDHVEHLGIFKXHLIPFLVPTXDKRWLSILTRQGLPELISLNCDSVRYTYRINEVTGETKDTNLAIMTSVOKLAQIQGIDLEWFSYFKDELITDESATLOSNEIKQIKTNKLAJDLKESQFALKYFYPHLKSIATGKS HDLI FDSVLKLSQKHSIQAPFVLCDSYVRNHSAEVDQGHALHALLSCLDLPDA KSRVLYLEKTVLSVMEDLMTLGGQVDSRMDGLMLRELMSLLKVPCTGHLEYPK GYLELGEIPNEQUPSNAVNTLHNDMPPEPVVTFVGMNDAEISNALTIFPQHGFD DMAKNVFLQDSVPYVDHPELSNLYHAYISFSVRNKPYSVLVLTFTETGRVPVADS PISFDAYRRCLESTK"

join(651..1816,1882..2003,2058..2116)

/gene="dmaW"

/allele="1"

/notes="aromatic prenilyltransferase; L-tryptophan dimethylallyl transferase; 4-(gamma,gamma)-dimethylallyltryptophan synthase; DMAT synthase; 4-(3-methylbut-2-enyl)-L-tryptophan synthase; dimethylallyl-pyrophosphate; most likely true start codon might be the short isoform based on comparison with GenBank accession number L39640; both long and short isoforms give rise to functional protein as confirmed by heterologous expression in yeast"

/codon_start=1

/evidence=experimental

/product="dimethylallyltryptophan synthase short isoform"

/protein_id="AAP81210.1"

/db_xref="GI:32402655"

/translation="MSTAKDPNGVVEILSLIFDPPSNEORLWHSSTAPFAAMLDNA GYNIHQYRHGLIFKXHLIPFLGVPTXDKRWLSILTRQGLPELSLNCDSVRYTYRINEVTGETKDTNLAIMTSVOKLAQIQGIDLEWFSYFKDELITDESATLOS NEIYKQIKTNKLAJDLKESQFALKYFYPHLKSIATGKS HDLI FDSVLKLSQKHD SIQAPFVLCDSYVRNHSAEVDQGHALHALLSCLDLPKASRVKYLLEKTVLSV MEDLTQGVQDSRMDGLMLRELMSLLKVPCTGHLEYPKLGELISLPNEQDPSMA NYTLHNDMPPEPVVTFVGMNDAEISNALTIFPQHGFDMAKNVFLQDSVPYH DPELSNLYHAYISFSVRNKPYSVLVLTFTETGRVPVADS PISFDAYRRCLESTK"

CDS

1041 CAGATTCAAGCGGCTATGAGCTTGGAGTGGTTTACTTTCAAGATGAGCTGACGCTG 1100

QY 477 TCGAGGAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGCTCTTCAAGACA 536

Db 1101 GACGAATCGGAATCGGCCACACTTCAAAGTAATGAGCTGTGCTCAAGGAGCAGATAAGACG 1160

QY 537 CAGAACAACTGGCAGCGGATCTGGAGCCATCTGGCCATATTGTCTTGAAGACCTACATC 596

Db 1161 CAGAACAGTTAGCTTGGATCTCAA---AGAAAGCCAGTTGGCGTCAAGTTACTTC 1217

QY 597 TACCCGGGATCAAGTCGATCGGACCGGACCCAAAAGAGAGACTCATGTTTGA---C 653

Db 1218 TATCCGCATCTTAAATCGATCGGACCGGCAAAATCCACACAGCATCTGATCTTCGACTCC 1277

QY 654 GCAATCAAGGCTGCGGACCAAGTTTGGCAAAAGTTGCCACTCCACTGGCAATCTCTCGAGGAG 713

Db 1278 GTGTTGAAGCTGTCCGAGAAGCATGACGATACAGCCCGCTTCCAGGTATTGTGCGCAC 1337

QY 714 TTTAT-----AGCTGAGCGAGCACCCACCTCTCTCGGCCAC 749

Db 1338 TATGTTTCGGCGGAAATCATTCGCGAAGTAGACCAACACGAGGCCCTACATGCGCGA 1397

QY 750 TTTCTCTCATCGGATTTGGTCAAGCGCTCGAGTCCCGAATCAAGTCTACTGTATGGAA 809

Db 1398 CTCTTGTCATCGGATTTGATCGATCCCGCAAGTCTCGCGTCAAGATATACCTACTGGAG 1457

QY 810 CGCAGCTCGACCTGGCCTCCATCGAAGGTATTGGACTCTCAACGGGCGACCGAAGCAT 869

Db 1458 AAGACGGTCTCGTTTGTCCGTGATGAAGATCTGTGAGCGCTGGCGGCGCAACAGTCGAC 1517

QY 870 CCAGAGACATGATGTGTGATGCGCTGAGGAGCTGTGGAGCTATTGCCCCGTACG 929

Db 1518 GCATCCACCATGAGCGGCCCTTGACATGCTTCGGAGCTCTGGAGCCCTGTGAAAGTCCC 1577

QY 930 GAGGGTCTGTCTCCACTCGCCAACTGCTTTTACGAGCGCGGTACCTCACCGCAGGAGCAG 989

Db 1578 ACTGGCCACTTGGAGTATCCAAAGGATCTCGAATTTGGAGAAATTCGACAGGAGCAG 1637

QY 990 CTCCTCTTAAATAAATTTTACCTGTCTCTTAAAGCGCACTTCCCGAACACAGATC 1049

Db 1638 CTCCTATCCATGCGCAACTATACACTGCACCAACAGACCCCATGCTGAGCCCCAGGTG 1697

QY 1050 TATTTCCTCTGTTTGGGCGAGACGACAAACCATCGCGAAGGATTGGCCACTTCTTT 1109

Db 1698 TATTTTACGGTTTTCGGCATGAATGACGCGAATATAGCAATGCTTTGACCATCTTCTTC 1757

QY 1110 GAGAGCAGAGTTGGGGTGGCTTGGCTTAAGAGCTATCCAG 1149

Db 1758 CAACGTCACGGATTGACGACATGGCGGAAAACTACCGAG 1797

RESULT 6

CPU312753

LOCUS

DEFINITION

Claviceps purpurea cpd2 gene for putative

dimethyl-allyl-tryptophan-synthase, exons 1-3, strain T5.

AJ312753

ACCESSION

AJ312753.1

VERSION

cpd2 gene; dimethyl-allyl-tryptophan-synthase.

KEYWORDS

Claviceps purpurea

SOURCE

Claviceps purpurea

ORGANISM

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.

REFERENCE

1

Arntz, C. and Tudzynski, P.

Molecular analysis of dimethyl-allyl-tryptophan-synthase genes

Unpublished

2 (bases 1 to 2694)

Tudzynski, P.

Direct Submission

Submitted (20-APR-2001) Tudzynski P., Institut fuer Botanik, Westfaelische Wilhelms-Universitaet, Schlossgarten 3, 48149 Muenster, GERMANY

QY	477	TCGGAGGAAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGCTTTCACAGACA	536
DB	1351	GACGAATCTGAATCGGCCATACCTTCAAGATACTAGCTGGTCAAGAGCAGATAAAGACG	1410
QY	537	CAGAAACAACTGGCAGCCGATCTGGAGCCATCTGGCGATATTGTCTTGAAGACCTACATC	596
DB	1411	CAGAACAAAGTTAGCCTTGGATCTCAA---AGAAAGCCAGTTTCGGCTCAAAGTTTACTTC	1467
QY	597	TACCCGCGGATCAAGTCGATCGAGCCGGACCCGAAAGAGAGACTCATGTTTGAC---	653
DB	1468	TATCCGCATCTCAAATCGATCGGACCGGCAATTTCCACACACTTTCTCATCTTCGACTCT	1527
QY	654	GCAATCAAGGCTGCCGACCAAGTTTG--GCAAGTTGGCACTCCACTGGCAATCCTCGAG	710
DB	1528	GTGTTCAAGCTGTCGCAAGCATGACAGCATACAGCCGGTTCAGGCAATTTGCGAC	1587
QY	711	GAGTTT-----ATAGCTGAGGAGCACCCACCTCTCTCGGCCAC	749
DB	1588	TATGTTTCGGCGGAAATGATTCCTCAGAAGTAGACCAACACAGAGCCCTACATGCGCGA	1647
QY	750	TTTCTCTCATCGATTTGGTCAAGCCGTCGAGTCCCGAATCAAGGTCTACTGTATGAA	809
DB	1648	CTCTTGTATCGGATTTGATCGATCCCGCCAAGTCTCCGTCAGAATATACCTCGAGGAG	1707
QY	810	CGCCAGCTCGACTCGGCTCCATCGAAGTATTTTGGACTCTCAACGGCGCAGCAACGAT	869
DB	1708	CAGACGGTCTCATTTGCCCGCGATGGAAGATCTGTGGAAGCTGGGCGCGCAGAGTCGAC	1767
QY	870	CCAGAGACACTGGATGCTCGATCGCTGAGGAGCTGTGCGAGCTATTGCCGCTCAGC	929
DB	1768	GCATCCACCATGAGCGGCTTGACATGCTTCGCGAGCTCTCGAGCCTGCTGAAAGTTCCC	1827
QY	930	GAGGTCCTGTGTCCTCCAGCTGCCAACTGCTTTTACGAGCGGGTACTCTCACCGCAGGAGCAG	989
DB	1828	ACTGCCACTTGGAGTATCCAAAGGATATATGGAATTTGGGAGAAATTCCTTAACGAGCAG	1887
QY	990	CTCCCTTCATTATAAATTTTACCTTGTCTCTTAAAGCGCACTTCCGGAACCAACAGATC	1049
DB	1888	CTTCCATCCCTGGTCAACTACACACTGCACCGCAACGACCCCATGCTCTGAACCCAGGTG	1947
QY	1050	TATTTCCCTGCTTTTGGCGAGAACACAAACCATTCGCGGAAGGATTGGCCACCTTCTTT	1109
DB	1948	TATTTCAAGTTCGGCATGATGACGCGCAATCAGCAATGCTTTTGACCAATTTTCTC	2007
QY	1110	GAGAGCAGAGGTTGGGTTGGCTTGGCTAAGAGCTATC	1146
DB	2008	CAACGTCAGGATTTTCCGACATGCGCAAAAGTACC	2044
RESULT 7			
AR181512			
LOCUS			
DEFINITION			
Sequence 1 from patent US 6335188.			
ACCESSION			
AR181512			
VERSION			
AR181512.1			
GI:20223726			
KEYWORDS			
SOURCE			
Unknown.			
ORGANISM			
Unknown.			
REFERENCE			
1 (bases 1 to 1347)			
AUTHORS			
Schardl, C.L. and Wang, J.			
TITLE			
Endophyte ergot alkaloid synthetic compounds, compounds which			
encode thereof and related methods			
JOURNAL			
Patent: US 6335188-A 1 01-JAN-2002;			
FEATURES			
Location/Qualifiers			
source			
1..1347			
/organism="unknown"			
/mol_type="unassigned DNA"			
ORIGIN			
Query Match 14.7%; Score 205.2; DB 6; Length 1347;			
Best Local Similarity 50.8%; Pred. No. 1.2e-49;			
Matches 647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;			

QY 33 GCAGCAACACTGCTGCGCAAGCCCTTCTACGTCTGAGTCAAGCCCTGACCTCTCGAAC 92
Db |||||
10 GCAAGACACTCCACAGGAGTTTATCATACCTTTAGCGAAACATTTGACTTTGCCAAC 69
QY 93 AAGAACACACAAATATGGTGGTATAGCAAGCTCCGATGTTTCCACCAATGATGCGGGG 152
Db |||||
70 AATGACAGAGGCTATGGTGGCACAGCAGCGGCCAAATGTTGAAAGATGCTCCAAACT 129
QY 153 GCGGCTATGATGTTACGCGACAGTCAAGTTCCTCTGTATCCACCGTGAGGTTCATC 212
Db |||||
130 GCTAACTATAGCATTTGATCTGATTCGATCTGGCATTTTAAAGAGCATGTCAT 189
QY 213 CCGGCGTTGGGTCATACCCAGAAAGGTCAGCCCATGCACTGGAAGATGCTCTCACA 272
Db |||||
190 CTTTCTTGGTGTATCTCTACAAAGTGG---CGAGCGGTGGCTAAGCATTTCTTAGC 246
QY 840 ATTTGAGCTCTCAACGGGGACCGAAGCATCCAGAGACACTGGATGGTCTGGATGCGGTG 899
Db |||||
844 CTTTGGACGCTTGGCGCGCTCGAGAAGATCAGTCCACTATTGAGGATGAGGATGATC 903
QY 900 AGGAGCTGTGGAGCTATTTGCCGTTCAGAGGGTCTGTGCTCACTGCGCACTGCTT 959
Db |||||
904 CGAGAACTTTGGGGTCTCTCAACATGCTCTCGGTGTTGCGCGCTTACCTGAGCCTTAC 963
QY 960 TAGAGCGGGTACTCCAGCGAGGAGCTCCCTCTCATTTAAATTTTACCTTTGCT 1019
Db |||||
964 TTGCGCTTGGCGCGATTTCCCATGAGCAACTTCGTCCATGCCCAATTTACACCTTACAC 1023
QY 1020 CTTAAAGGCACTTCCGAAACACAGATCTATTTCCCTGCTTTTGGGCGAGAACACAAA 1079
Db |||||
1024 CATAATGATCCCATAGCAACCGCAAGTGTACTTTTACTGTGTTGCGCATGAATGATG 1083

QY 1080 ACCATCGGGAAGGATTGGCCACCTTCTTTGAGACGAGGTTGGGTGGCTTTGGCTAAG 1139
Db |||||
1084 GAGGTGACTAATGCACTCAGACATTTCTCATGAGGATGAATGAGCGATATGGCAAGT 1143
QY 1140 AGCTATCCAGCGGATTTGGCATCTACTATCCGATGCGACCTGCGAGCGCAATCAC 1199
Db |||||
1144 AATACAAAGCCTCCCTCAGGGAATTTTCCGCGATCATGATTACGAAGCCCTGAATAT 1203
QY 1200 CTGAGGCGGTGATCTCCTCTCTTACAAGGGGAAACCGTACATGATGATGTTGATCCTC 1259
Db |||||
1204 ATCCACTCGTACATTTCTCTCTCTCCGAAAGAACAGCCATATTTAAGTGTGATCTC 1263
QY 1260 CATACCTTCGAGC 1273
Db |||||
1264 CACTCCTTTGAAC 1277

RESULT 8
AR181513
LOCUS AR181513 1353 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 3 from patent US 6335188.
ACCESSION AR181513
VERSION AR181513.1 GI:20223727
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1353)
AUTHORS Schardt, C.L. and Wang, J.
TITLE Endophyte excret alkaloid synthetic compounds, compounds which encode therof and related methods
JOURNAL Patent: US 6335188-A 3 01-JAN-2002;
FEATURES
source location/Qualifiers
1..1353
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 14.7%; Score 205.2; DB 6; Length 1353;
Best Local Similarity 50.8%; Pred. No. 1.2e-49;
Matches 647; Conservative 0; Mismatches 598; Indels 39; Gaps 5;

QY 33 GCAGCAACACTGCTGCGCAAGCCCTTCTACGTCTGAGTCAAGCCCTGACCTCTCGAAC 92
Db |||||
10 GCAAGACACTCCACAGGAGTTTATCATACCTTTAGCGAAACATTTGACTTTGCCAAC 69
QY 93 AAGAACACACAAATATGGTGGTATAGCAAGCTCCGATGTTTCCACCAATGATGCGGGG 152
Db |||||
70 AATGACAGAGGCTATGGTGGCACAGCAGCGGCCAAATGTTCCAAAGATGATCTCCAAACT 139
QY 153 GCGGCTATGATGTTACGCGACAGTCAAGTTCCTCTGTATCCACCGTGAGGTTCATC 212
Db |||||
130 GCTAACTATAGCATTTTATGCTCAGTATCAACATCTGAGCATTTTAAAGAGCCATATCAT 189
QY 213 CCGGCGTTGGGTCATACCCAGAAAGGTCAGCCCATGCACTGGAAGATGCTCTCACA 272
Db |||||
190 CTTTCTTGGTGTATCTCTACAAAGTGG---CGAGCGGTGGCTAAGCATTTCTTAGC 246
QY 273 CGCTTCGAGCTTCTCTTCGAGCTGAGCTTCAATTTACTCCAAATCACTACTACCGTTTGA 332
Db |||||
247 AGATACGGAACCGGTTGAGCTAAGTCTTAAATGCTCTGACTCCATAGTTTGGTATACA 306
QY 333 TTGAGCGCCCTCGGTTCCCTGACGGGAACGAAAGATGATCCATTTCAACCCAGGCAATC 392
Db |||||
307 TACGAGCCTATTAAACGCGCAACTGGCAGCCATCTGGATCCGTTTCAACATTTTCGCTATC 366
QY 393 AGGCGCTGCTCCAGGACCTCAAGGCCATGTTTCCAGGGCTTCACCTGGAATGTTTCGAT 452
Db |||||
367 TGGAGGCTTTAAAGAGCTTATAGATTCAGCCAGGAGATAGCTTCAATGGTTTTC 426
QY 453 CATTTCACTAAAGCATTTGGTGGTTTCGGAGGAAGAGCTCGACTCTGTAGATCGAGAT 512
Db |||||
427 TACTTTAAACAGAGCTTACCTTGACGCAACAGTCCACGTACCTGCACTCGCAAAAC 486
QY 513 ATTGAGATCCCGCTTTCAAGACACAGAACAACTGGCAGCGCATCTGGAGCCATCTGGC 572
Db |||||
487 TTGGTTAAGGAACAGATCAAACTCAAAACAAAGCTGCTTTGAGCCTTAA---AGTGCAC 543
QY 573 GATATTGCTTTGAGACCTACATCTACCGCGGATCAAGTCCGACCGCCGCGGCCCA 632
Db |||||
544 AAGTTCGTACTGAGACCTACATCTACCGCGATGAGTCCGTGCAACTGTTAAATCG 603
QY 633 AAGAGAGACTCATGTTTGAAGCAAT---CAAGGCTGCGGACAAAGTTTGGCAAGTTCGC 689
Db |||||
604 GTCCAGAGCTGCTGTTGGCTCCGTCGCGCAAGCTAGCGCAGAACACAAAGATATCCGT 663
QY 690 ACTCCACTGCAATCTCGAGGAGTTTAT---AGCTGAGCGAGCACCCACCTCTCCG--- 745
Db |||||
664 CTTGCTTTGAAATGCTAGAAGACTATGTCAGTCTCGCAATAAATTTCTTACCAGGAT 723
QY 746 -----CCACTTTCTCTCATGCGATTGCTCAAGCGCTCC 779
Db |||||
724 GACAGTCACAATACTCTATTATCTTCACGCTTCTCTCTTGGCACTGATAAGTCTTACC 783
QY 780 GAGTCCGGAATCAAGTCTACTGTATGGAAGCGCAGCTCGACTGGCTCTCATCGAAGT 839
Db |||||
784 AAGTCTGCTGTTCAAGATCTACCTCCCTGGAACGAATGGTCTGTTGCGCAGCATGGAAGAT 843
QY 840 ATTTGAGCTCTCAACGGGGACCGAAGCATCCAGAGACACTGGATGGTCTGGATGCGGTG 899
Db |||||
844 CTTTGGACGCTTGGCGCGCTCGAGAAGATCAGTCCACTATTGAGGATGAGGATGATC 903
QY 900 AGGAGCTGTGGAGCTATTTGCCGTTCAGAGGGTCTGTGCTCACTGCGCACTGCTT 959
Db |||||
904 CGAGAACTTTGGGGTCTCTCAACATGCTCTCGGTGTTGCGCGCTTACCTGAGCCTTAC 963
QY 960 TAGAGCGGGTACTCCAGCGAGGAGCTCCCTCTCATTTAAATTTTACCTTTGCT 1019
Db |||||
964 TTGCGCTTGGCGCGATTTCCCATGAGCAACTTCGTCCATGCCCAATTTACACCTTACAC 1023
QY 1020 CTTAAAGGCACTTCCGAAACACAGATCTATTTCCCTGCTTTTGGGCGAGAACACAAA 1079
Db |||||
1024 CATAATGATCCCATAGCAACCGCAAGTGTACTTTTACTGTGTTGCGCATGAATGATG 1083

QY	416	GGCCATGGTTCAGGGCTTGACCTGGAATGGTTGATCATTTTCACTAAAGCATTTGGTGGT	475
Db	729	TCGGATTTCAGGCAAGTATTGATCTTGAGTGGTTACCTACTTTCAAAGATGAGCTCACATGT	788
QY	476	TTCCGGAGGAAGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGTCTTCAAGAC	535
Db	789	CAATATGACGGAGTCTGAATATCTCGGAGTAAATGGTTGGTGAATCTCAGATAAAGAC	848
QY	536	ACAGAAACAACTCGCAGCCGATCTGAGCCATCTGGCGATATGTCTTGAAGACCTACAT	595
Db	849	CCAGAAACAACTGGCTTGGATCTCAAG--GGGGATCAATTCACCTAAAAGTCTATAT	905
QY	596	CTACCCGGATCAAGTCCGATCGGACCGGACCCCAAGAGAGACTCATCTTTGACGC	655
Db	906	GTATCCGAGTTAAGTTCGGTTGCAACCGGCAATCATTTACATGAGCTTATTTTGGATC	965
QY	656	---AATCAAGGCTGCGACAAAGTTTGGCAAAAGTTGCCACTCCACTGGCAATCTTCGAGCA	712
Db	966	GGTACGAGGCTGTCACTGAAGTATAACAGTATTCGTGCACCGCTCGACATGTTGGACGA	1025
QY	713	GTATTAGCTGAGC-----AGCACCCACCTCTCGGCCA	748
Db	1026	TTACGTTACATCCGGAACATGGGAGCCAAATGGCGAAGAACACCTTTCCTTTGGACCCCG	1085
QY	749	CTTTCTCTCATGCGATTGTGTCAGCGCTCCGAGTCCCGAATCAAGGTCTACTGTATGGA	808
Db	1086	TCCTCTCTTGTGATTAAATGATCCTCATAGTCTCGTGTCAAGATATCTTGCACGA	1145
QY	809	ACGCCAGCTGACCTGCTCCATCGAAGTATTTGGACTCTCAACGGCGGACGGAACGA	868
Db	1146	CGGAAAGGTCTCATTTGTCGCGATGGAAGATCTATGGACACTCGGCGGACGACGACGA	1205
QY	869	TCAGAGACACTGATGGTCTGATCGCTGAGGAGCTGTGGACACTATTCGCGTAC	928
Db	1206	TTCAACCAACATGATGGGCTTAAATGTTTCGGAACATATGGATCTGCTTGAATACC	1265
QY	929	GGAGGCTCTGTCTCACTGCGCACTGCTTTTACGAGCCGGTACCTCACCGAGGACGA	988
Db	1266	GACTTGTCTTACAGAGTACCTCGACCACTTCTTAGAGCTGGGTGAGATTCGAAGAGACGA	1325
QY	989	GCTCCCTCTTATTAATTTTACCTTGCTCTCTAAAGCGGACTTCCCGAACCCACAGAT	1048
Db	1326	ACTTCCATCTATGGTAACTATACATATGCAATGGCGGATCTTATGCCCGAACCAAGT	1385
QY	1049	CTATTTCCCTGCTTTTGGGCGAGAACACAAACCATCGCGGAGGATTTGGCCACCTTTT	1108
Db	1386	GTATTTACAGTGTTTGGCATGAATGACTCCAAGGTGATCAGTGCATTAAACGGAATTTT	1445
QY	1109	TGAGACAGAGGTTGGGTGGCTTGGTAAAGACTATCCAGCGGATTTGGCATCTCTACTA	1168
Db	1446	CAAGCGCGGTGTGGAAATGGCAATGCAAGAAATACAGAGCCTTTTTCAAAACACTCAT	1505
RESULT 10			
CPU312752			
LOCUS	2570 bp	DNA	linear
DEFINITION	Claviceps purpurea partial cpd2 pseudogene for putative dimethyl-allyl-tryptophan-synthase, exons 1-3.		
ACCESSION	AJ312752		
VERSION	AJ312752.1 GI:13810196		
KEYWORDS	cpd2 pseudogene; dimethyl-allyl-tryptophan-synthase.		
SOURCE	Claviceps purpurea (ergot fungus)		
ORGANISM	Claviceps purpurea		
REFERENCE	1		
AUTHORS	Arentz, C. and Tudzynski, P.		
TITLE	Molecular analysis of dimethyl-allyl-tryptophan-synthase-genes		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2570)		
AUTHORS	Tudzynski, P.		

JOURNAL	Submitted (20-APR-2001) Tudzynski P., Institut fuer Botanik, Westfaelische Wilhelms-Universitaet, Schlossgarten 3, 48149 Muenster, GERMANY		
FEATURES	Location/Qualifiers		
source	1..2570		
gene	/organism="Claviceps purpurea" /mol_type="genomic DNA" /db_xref="taxon:5111" 901..2370 /gene="cpd2" /pseudo		
CDS	join(901..2071,2137..2257,2312..2370) /gene="cpd2" /functions="isoprenylation of tryptophan" /note="7 bp insertion (1036-1042) results in the first exon being out of frame" /pseudo		
exon	/codon_start=1 /product="putative dimethyl-allyl-tryptophan-synthase" 901..2071 /gene="cpd2" /number=1 /pseudo		
intron	2072..2136 /gene="cpd2" /number=1 2137..2257 /gene="cpd2" /number=2 /pseudo		
exon	2258..2311 /gene="cpd2" /number=2 2312..2370 /gene="cpd2" /number=3 /pseudo		
intron	2312..2370 /gene="cpd2" /number=3 /pseudo		
exon	2312..2370 /gene="cpd2" /number=3 /pseudo		
ORIGIN	Query Match 13.9%; Score 194; DB 8; Length 2570; Best Local Similarity 52.1%; Pred. No. 3e-46; Matches 579; Conservative 0; Mismatches 495; Indels 38; Gaps 5;		
QY	67	TGAGTCAAGCCCTGAACCTCTCGAACAGGACACACACAAAATGGTGTATAGCACAGCTC	126
Db	944	TGAATCTGATTTTGGAGTTCCTCCAGCAATGAACGACTATGGTGCCACAGTACCGCGC	1003
QY	127	CGATGTTTGCACCATGATGGCGGGCCGGCTATGATGTTTCACGCACAGTCAAGTTC-	185
Db	1004	CTATGTTTGGCGGATCCTTGACAAATCGCGCTACACTGTACAGTCCACGACCAATATCG	1063
QY	186	----CTCTGATCCACGCTGAGGTCAATCCCGGGCTTGGTCCATACCCAGAAAGGG	241
Db	1064	GCATCTGAGCATTTTCAAGACACACATTTATCCCTTTTCTTGGTGTCTATCCAAACAAAGA	1123
QY	242	TCAGCCCATGCACTGGGAAGAGTCACTCACACGCTTCGGACTTCCTTTTCGAGCTGAGCTT	301
Db	1124	CAAGGAAGA---TGGCTCAGCATCTCCACAGATGGGTCTTCTTTTGAACGTAGTTT	1180
QY	302	CAATTACTCCAAATCACTACTACGGTTTGGATTCGAGCCCTCGTTCCTCTGACGGGAAC	361
Db	1181	GAATTGACCGACTCTCTGTTTCGATATACATAGAGCCCATCAATGAGATGACGGGGAC	1240
QY	362	GAAGATGATCCATTCAACACCCAGGCAATCAGSCCTGTTTCTCAGGACCTCAAGGCCAT	421
Db	1241	GGAAAGATCCATCAATACGTTGCGATATAGGAAGTGTCCAAAGCTTGCCCATAT	1300
QY	422	GGTTCCAGGGCTTCACTCGAATGGTTCGATCAATTCATTAAGCATTTGCTGTTTCGGA	481
Db	1301	TCAAGCGGTATCGACTTGGATGGTGGTTAGTTACTTCAAGGATGAGTTGACGTTGGACGA	1360
QY	482	GGAGAGGCTCGGACTCTCTAGATTCGATATTTAGATCCCGCTTCTTCAAGACACAGAA	541

Db 1361 ATCGGATCGGCCACTACTTCAAGATAGTCTGAGCTGGTCAAGGACGACATMAAGACGAGAA 1420
Qy 542 CAAACTGGCAGCCGATCTGGAGCCATCTGGCGATATTTGTCTTGAAGACCTTACATCTACCC 601
Db 1421 CAAAGTAGCCTTGGATCTCAA---AGAACCCAGTTTCGGCTCAAAGTTTACTTCTATCC 1477
Qy 602 GCGGATCAAGCTCGATCGCGACCGGACCCCAAAGAGAGACTCATGTTTACGCG----- 655
Db 1478 GCATCTCAAAATCGATCGCGACCGGCAATCCACACACTTTTCTCATCTTTGACTCCGTGTT 1537
Qy 656 -----AATCAAGGCTGCCACAAAGTTTGGCAAAGTTTGGCAAGTTGCACTCCACTGGCAA 702
Db 1538 CAAAGTTGTCGAGAAGCATGACAGCATACAGCCGGTTCACGAGCATTTGGGACTATGT 1597
Qy 703 TCTTCGAGGAGTTTATAGCTAGGAGCACCCCA-----CCCTCTCGGCCACTTTCT 754
Db 1598 TTCCGCGCGAAATTAATTCGACAGATGAGACCAACACGAGCCCTACATGCGCGTCTCT 1657
Qy 755 CTATGCGATTTGGTCAAGCGCTCCGAGTCCCGAATCAAGGTCTACTGTATGGAAGCCCA 814
Db 1658 GTATGCGATTTGATCGATCCCGCAAGTCTCGGTCAAGATATACTGCGAGGAGCAGAC 1717
Qy 815 GCTCGACTGGCTCCATCGAAGTATTTGGATCTCAACGGCGACCGGAACGATCCAGA 874
Db 1718 GGTCTCATTTGTCGCGATGGAAGATCTGTGGAGCTGGGGCGCGCAAGTCCGACGCATC 1777
Qy 875 GACACTGGATGTTGATCGGCTGAGGAGCTGTGCGAGCTATTGCCGCTCAGCGAGGG 934
Db 1778 CACCATGGATGGCTTGAATTTCTTCGCGAGCTCTGGAGCTGCTAAAGTTCCCGCTGG 1837
Qy 935 TCTGTGTCCACTGCGCACTGCTTTTACGAGCGGGTACCTCACCGCAGGAGCAGCTCCC 994
Db 1838 CCATCTGGAGTATCCGGAAGATATATGGAATTTGGGAGAAATTTCCCAACGAGCAGCTCC 1897
Qy 995 CTTCAATTAATTTTACCTTGTCTCTTAAAGGCACTTCCGACACACAGATCTATT 1054
Db 1898 ATCCCTGGTCAACTACACTGACACCAACGACCCCATCGCTGAACCTCAGGTGTATTT 1957
Qy 1055 CCTGTCTTTGGGAGAACCAACCAATCGCGAGGATTTGCCACCTTCTTTGAGAG 1114
Db 1958 CACGTTTTGGCATGATGACGCCGAAATCAGCAATGCTTTGACATCTTCTTCCACG 2017
Qy 1115 CAGAGTTGGGTGGCTTGGCTAAGAGCTATC 1146
Db 2018 TCAGGATTTGACACATGCGGAAAAAGTACC 2049

RESULT 11
AY262014 1783 bp DNA linear PLN 12-DEC-2003
LOCUS Claviceps purpurea dimethylallyltryptophan synthase (dmaW) gene,
DEFINITION dmaW-2 allele, partial sequence.
ACCESSION AY262014
VERSION AY262014.1 GI:32967583
KEYWORDS Claviceps purpurea (ergot fungus)
SOURCE Claviceps purpurea
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
REFERENCE Wang,J., Machado,C. and Scharld,C.L.
AUTHORS The determinant step in ergot alkaloid biosynthesis by a grass
TITLE endophyte
JOURNAL Fungal Genet. Biol. (2004) In press
REFERENCE 2 (bases 1 to 1783)
AUTHORS Wang,J., Machado,C. and Scharld,C.L.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2003) Plant Pathology, University of Kentucky,
201 Plant Science Building, Lexington, KY 40546-0312, USA
FEATURES
1. .1783
/organism="Claviceps purpurea"
/mol_type="genomic DNA"

/isolate="ATCC 20102"
/specific_host="Secale cereale"
/db_xref="ATCC:20102"
/db_xref="taxon:5111"
140. .>1609
/gene="dmaW"
/allele="2"
/note="dimethylallyltryptophan synthase; null allele due to frameshift"
275. .281
/gene="dmaW"
/allele="2"
/note="insertion; probable transposon footprint"

ORIGIN
Query Match 13.9%; Score 193.6; DB 8; Length 1783;
Best Local Similarity 52.11%; Pred. No. 3.8e-46;
Matches 580; Conservative 0; Mismatches 46; Indels 40; Gaps 5;

Qy 67 TGAGTCAAGCCCTGAACTCTCGAACAGGACACACAAAATGTGGTATAGCAGCTC 126
Db 183 TGATCTGATTTGGAGTTCCCGCATGTAACAGCGACTATGGTGGCAGTAGTACCGCGC 242
Qy 127 CGATGTTTGCACCATGATGGCGGGCGGCTATGATGTT-----CACGCACAGTAC 179
Db 243 CTATGTTTGGCGGATGCTTGACAATGCCGCTACACTGTACAGTGTCCACGACCAATAT 302
Qy 180 AGTTCTCTGTATCCACGCTGAGTGTATCATCCCGCGTTGGGTCCATACCCAGAAAAG 239
Db 303 CGGCATCTGAGCATTTTCAAGACACATTAATTCCTTTCTTGTGTGTATATCCACAAAA 362
Qy 240 GGTGAGCCATGCTGACTGGAAGAGTCTCTCACAGCTTCGGACTTCTCTTCGAGCTGAGC 299
Db 363 GACAAGGAAG---ATGGCTCAGCATCTCACCAGATCGGCTCTCTCTTGGAACTGAGT 419
Qy 300 TTCAATTAATCAATCACTACTACGGTTTGGATTCGAGCCCTCGGTTTCCCTGACGGGA 359
Db 420 TTGAATTTGATCCGACTCTGTGTTTGCATATATACAGAGCCCATCAATGAGATGACGGG 479
Qy 360 ACGAAGGATGATCCATTCACACCCAGCAATCAGGCGCTGTTCTCCAGAGCTCAAGGCC 419
Db 480 ACGGAGAAGATCCATTCATACGTTGCCGATATAGGAAGTGTCCAAAAGCTTCCCCAG 539
Qy 420 ATGGTTCCAGGCTTACCTGGAATGTTGATGATCAATTTCACTAAAGCATTTGTTGTTTCG 479
Db 540 ATTCAAGCGGGTATCGACTTGGAGTGGTTTACTTCAAGGATGAGTTCAGCTTGGAC 599
Qy 480 GAGGAAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGCTTTCAGAGCACAG 539
Db 600 GAATCGGATCGGCCATATCTTCAAGATACTAGCTGTGGTCAAGGAGCAGATAAAGACGCG 659
Qy 540 AAACAACTGGCAGCCGATCTGGAGCCATCTGGCGATATTGTTTGAAGACCTTACATCTAC 599
Db 660 AACAAAGTTAGCTTGGATCTCAA---AGAAGCCAGTTCCGCTCAAAGTTTACTTCTAT 716
Qy 600 CCGCGGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 655
Db 717 CCGCATCTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 776
Qy 656 -----AATCAAGGCTGCGACAAAGTTTGGCAAGTTTGGCAAGTTTGGCAAGTTTGGCA 700
Db 777 TTCAAGTTGTTCGACAGACGATGACAGCATACAGCCCGCTTCCAGGCAATTTGGCACTAT 836
Qy 701 AATCTCGAGGAGTTTATAGTGAAGCAGACCCCA-----CCCTCTCGGCCACTTTT 752
Db 837 GTTTCGCGCGAAATTAATTCGCGAGATGATAGACCAACACGAGCCCTACATGCGCGTCTC 896
Qy 753 CTCTCATGAGTTTGGTCAAGCGCTCCGAGTCCCGAATCAAGGTTTACTGTATGGAAGCG 812
Db 897 TTCTCATGCGATTTGATGATCCCGCAAGTCTCCGCTCAAGATATATCTTCGAGGAGCAG 956
Qy 813 CAGCTCGACCTGGCTCCATCGAAGTATTTGGACTCTCAACGGCGGACGGAACCATCCA 872

Db 957 ACGGTCTCATTTGTCGCGATGGAAGATCTGTGGAACGCTGGCGCGCGACAAGTCGACGCA 1016
 Qy 873 GAGACACTGGATGCTCGGATCGCTGAGGAGCTGTGCGAGCTATTGCCCGTCACGGAG 932
 Db 1017 TCCACCATGGATGCCCTTGAATCTTCGCGAGCTCTGAGAGCTGTAAAGTTCGCGCT 1076
 Qy 933 GGTCTGTGTCACATGCGCACTGCTTTTACGAGCGGATACCTCACGCGAGGAGAGCTC 992
 Db 1077 GGCACCTTGGAGTATCCGGAAGGATATGGAATTTGGAGAAATTCACACGAGCAGCTT 1136
 Qy 993 CCCTTCATTATAATTTACCTTGTCTCTTAAAGGCGACTTCCCGAACACACAGATCTAT 1052
 Db 1137 CCATCCCTGTCACTACACTGACACACAGACCCCATGCTGAACTCAGGTGTAT 1196
 Qy 1053 TTCCCTGCTTTTGGCGAGAACGACAAACCATCGCGAAGGATGGCCACCTTCTTTGAG 1112
 Db 1197 TTACGGTTTTCGGCATGATGACGCGAAATCAGCAATGCTTTGACCATCTTCTTCCAA 1256
 Qy 1113 AGCAGAGTTGGGTGCTTGGCTAAGAGCTATC 1146
 Db 1257 CGTCACGATTTGACGACATGGCGAAAAAGTACC 1290

RESULT 12
 CLCDMAW 1894 bp DNA linear PLN 04-JUN-1998
 LOCUS Claviceps fusiformis dimethylallyl-trans-transferase (dmaW) gene,
 DEFINITION complete cds.
 ACCESSION L39640.1 GI:1005417
 VERSION 1
 KEYWORDS dmaW synthase; dimethylallyl-trans-transferase; dimethylallyldiphosphate; dimethylallyltransferase; dmaW gene; prenyltransferase; tryptophan.
 SOURCE Claviceps fusiformis
 ORGANISM Claviceps fusiformis
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
 REFERENCE 1 (bases 1 to 1894)
 AUTHORS Tsai H.F., Wang H., Gebler J.C., Poulter C.D. and Schardl C.L.
 TITLE The Claviceps purpurea gene encoding dimethylallyltryptophan synthase, the committed step for ergot alkaloid biosynthesis
 JOURNAL Biochem. Biophys. Res. Commun. 216 (1), 119-125 (1995)
 MEDLINE 96067540
 PUBMED 7488077
 FEATURES
 Location/Qualifiers
 1..1894
 /organism="Claviceps fusiformis"
 /mol_type="genomic DNA"
 /db_xref="taxon:40602"
 /clone_lib="H-F Tsai's pMOcosX"
 /dev_stage="mycelium"
 /tissue_lib="ATCC 26245"
 161..1894
 /gene="dmaW"
 join(161..1381,1456..1577,1640..1894)
 /gene="dmaW"
 /product="dimethylallyl-trans-transferase"
 161..1381
 /gene="dmaW"
 /number=1
 /evidence=experimental
 join(195..1381,1456..1577,1640..1894)
 /gene="dmaW"
 /EC_number="2.5.1.1"
 /standard_name="DMAT synthase"
 /codon_start=1
 /product="dimethylallyl-trans-transferase"
 /protein_id="AAC1893.1"
 /db_xref="GI:1005418"
 /translation="MMTKAPATAVYDTLSLLDFPNQERLWHSIAPIVFAMLDITAG
 HNVDDYRLHGIKGIIFPLGVPAGKHPTWPSVLTRYGIFPELSLNCDSVVRYP
 EPTEHTGDDSYNAFLEICTQKLVRIQPGIDMEWFSYFRNELVLNATESLGRN
 DSVNQPIRTONKLAIDLKDRFALKVLYLPHLKSITATGVSSHDLIFNSVRKLSQHT

Query Match 13.1%; Score 182.6; DB 8; Length 1894;
 Best Local Similarity 51.5%; Pred. No. 7.6e-43;
 Matches 586; Conservative 0; Mismatches 494; Indels 57; Gaps 5;
 Qy 61 ACGTGTGAGTCAAGCCCTGAACCTCTCGAAGAGGACACACAAAATGGTGTATAGCA 120
 Db 229 ACACCTGAGTCTGCTTTTTCATTTCCCAATCAAGAGACGCGACTTTGGTGGCAGTA 288
 Qy 121 CAGCTCCGATGTTGCCACCATGATGGGGGGCGGCTATGATGTTTCACGACAGTACA 180
 Db 289 TCGGCCCATGTTTGGCGCCATGCTCGATCTGTGACACACAGCTTCACGACCAATATC 348
 Qy 181 AGTTCTCTGTATCCACCGTGAGTCTATCATCCCGGGCTTGGGTCCATACCCAGAAAAG 240
 Db 349 GACACCTGGGCATTTTCAAGAAGCACATCATCCCTTTCTCGGGGGTCTATCCAGCGCAAG 408
 Qy 241 GTCAGCCCATGCACTGGAAGAGTCATCTCACAGCTTCGGACTTCCTTTTCGAGCTGAGCT 300
 Db 409 GAAAGCACAC---CATGGCCAGTGTCTCTACCCAGATATGGCATACCGTTTGAAGTAC 465
 Qy 301 TCAATTACTCCAAATCACTACTACGGTTTGCATTTCGAGCCCTCGGTTCCTCTGACGGGAA 360
 Db 466 TCAACTGCTGACTCGCTTGTCCGGTACACATTCGAACCCACACTGAGCACACTGGCA 525
 Qy 361 CGAAGGATGATCCATTCAACACCCAGGCAATCAGGCGCTGTTCACAGGAGCTCAAGGCCA 420
 Db 526 CAGGTGATGATTCATACACGCAATTCGATTTCTGGAATGTATCCAAAAGCTTTGTCCGGA 585
 Qy 421 TGGTTCCAGGCTTACCTGGAATGTTTCGATCATTTTCAAGAGCATTCGTTTCGTTTCGG 480
 Db 586 TCCAGCCGGGAATCGACATGGATGAGTTCAGTACTTTCAGGATGAGTTCGTTTCTGACG 645
 Qy 481 AGGAAGAGGCTCGGAGCTCTGCTAGATCGAGATATTGAGATCCCC---GTCTCAAGACAC 537
 Db 646 CGACGGAATCTGCACGCTTTGGACCTAACGACTCGGTGAACCAACACCATACGAGACCC 705
 Qy 538 AGAACAACCTGGGAGCGGATCTGGGCCATCTGGCGATATTGCTTGAACCTCATCATCT 597
 Db 706 AGAACAAGCTGGCGCTTGGATCTCAAGGGCGATCG---TTTCGGCTCAAGGTCTATCTCT 762
 Qy 598 ACCCGGAGTCAAGTTCGATCGGACCGGACCCCAAGAGAGACTCATGTTTTCAC--- 653
 Db 763 ACCCTCATCTCAAGTCAATTCGACCGCGCTTTTCGTTCACACGACCTCATATTCAACTCGG 822
 Qy 654 -----GCAATCAAGCTCCGACAGAGTTTGGCAAGTT 686
 Db 823 TGGCCAAAGCTGTGCGAAGACACACTAGTAGTATTGAGCCCTCTTCAACGTTTGTGGACT 882
 Qy 687 GCCACTCACTGGCAATCTCTC-----GAGGAGTTTATAGCTGAGCGAG 729

ORIGIN

intron

exon

intron

exon

Db 883 ATGTGCGCTCGGAAACGATCCGATTGCAATGCAGCAGAAAGCAGAGGATTCAG 942
Qy 730 CACCCACCTCTCGGCCACTTTCTCTCATGCGATTGCTCAAGCGCTCGAGTCCGAA 789
Db 943 CCAGCGCTTACGAGCGGCACTCTTGTGCTGTGACTTGTGTCGATCGTCCAAATCTCGCA 1002
Qy 790 TCAGGCTTACTGTATGGAAGCCAGCTCGAGCTGSCCTCCATCGAAGTATTTGAGCTC 849
Db 1003 TCAAGATATACCTTGTGGAAGCAGACAGTCTCTTGACCGCAATGGAAGATCTGTGACGT 1062
Qy 850 TCAACGGGGACGGAACGATCCAGAGACACTGTGATGGTCTGGATGCGCTGAGGGAGCTGT 909
Db 1063 TGGGGGACGAGAACCGACTCTTCCACCCTGAACGGGCTTGACATGATGCGAGAGCTTT 1122
Qy 910 GGCAGCTATGCCCCGTCGAGGGGTCTGTGTCCACTGCGCACTGCGAAGCTCTTTACGAGCGG 969
Db 1123 GGCATCTTTCGAGATTCGCTCGGCTTCATGAAGTATCCAGATCTGACTTTGAAGCTGG 1182
Qy 970 GTACCTCACCGAGGAGAGCTCCCTTTCATTAATAATTTTACCTTGTCTCTTAAAGCG 1029
Db 1183 GCGAAGTCCCGACGAGAGCTTCGTCATGGTCCATTTATGCTTGATCCCGACCAAC 1242
Qy 1030 CACTTCCGGAACACAGATCTATTTCCCTGCTTTTGGGAGAACGACAAACCATCGCGG 1089
Db 1243 CAATGCCCGAGCCACAGGTGTATTTACAGATGTTTGGCATGAGTGACGCGGAATCACCA 1302
Qy 1090 AAGGATGSCCACTTCTTTGAGAGCAGAGTTTGGGTGCTTGGCTAAGAGCTATC 1146
Db 1303 ATGCTTGGCCACCTTCTTCAGCCGACACGATGGTATGAGATGGCAAAAAGTACC 1359

RESULT 13
AR181515
LOCUS AR181515 1598 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 6 from patent US 6335188.
ACCESSION AR181515
VERSION AR181515.1 GI:20223729
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS Schardi,C.L. and Wang,J.
TITLE Endophyte ergot alkaloid synthetic compounds, compounds which encode therefor and related methods
JOURNAL Patent: US 6335188-A 6 01-JAN-2002;
FEATURES Location/Qualifiers
source 1..1598
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 12.7%; Score 176.6; DB 6; Length 1598;
Best Local Similarity 50.7%; Pred. No. 4.7e-41;
Matches 583; Conservative 0; Mismatches 529; Indels 39; Gaps 5;
Qy 33 GCAGCAACACTGTGCGAAAGCCCTTCTACGTGCTGAGTCAAGCCCTGAACTCTCGAAC 92
Db 34 GCAAGACACTCCACCAGGAAGTTTATCAAAACCTCAGCGAAACATTTGACTTTGCCAAC 93
Qy 93 AAGGACACACAAATGGTGGTATAGACAGCTCCGAGTTTCCACCATGATGGCGGG 152
Db 94 AATGACCAAGGCTATGGTGGCAGACGCGCGCCCAATGTTCCAAAGATCTCCAACT 153
Qy 153 GCGGCTATGATGTTACGCGACAGTACAAGTTCCTCTGTATCCACCGTGAAGTCAATC 212
Db 154 GCTAACTATAGCATTTATGCTCAGTATCAACATCTGAGCACTTATAAAGCCATATCAIT 213
Qy 213 CCGCGTGGTTCATACCCAGAAAGGCTCAGCCCATGCACTGGAGAGTCACTCACA 272
Db 214 CTTTCTTGGTGTCTATCTCACAAGAAGTGG---CGAGCGGTGGCTAAGCATTTCTAG 270
Qy 273 CGCTTCGAGCTTCTCTTCGAGCTGAGCTTCAATTAATCTCCAAATCACTACTACGTTTGA 332

Db 271 AGATACGGAACCCGTTTGAGCTTAAGTCTTAATTTGCTCTGACTCCATAGTTGGTATACA 330
Qy 333 TTCAGGCCCTTCGGTTCCCTGACGGGAAACGAGAGATGATCCATTCACACCCAGGCAATC 392
Db 331 TACGAGCCTATTACGCGCGCAACTGCGAGCCATCTGGATCCGTTTCAACACTTTTCGCTATC 390
Qy 393 AGGCTCTTCTCCAGGACCTCAAGGCCATGTTTCCAGGGCTTGACCTGGAAATGTTGAT 452
Db 391 TGGAGGGCTCTTAAAAAAGCTTATAGATCCAGCGAGGCATAGACCTTCATGTTTTC 450
Qy 453 CATTTCACTAAAGCATTTGGTTCGGAGAGAGAGCTCGGACTCTGTCTAGATCGAGAT 512
Db 451 TACTTTAAACGAAGAGCTTACACTTGACGCAAAACGAGTCCACGTACCTGCACTCTCAAAAC 510
Qy 513 ATTGAGATCCCGCTTTCAGACACAGAACAACTGGCAGCCGATCTGGAGCCATCTGGC 572
Db 511 TTGTCGAAGGAACAGATCAAAACACTCAAAACAAGCTAGCTTAGACCTTAA---AGGTGAC 567
Qy 573 GATATTGCTTGAAGAGCTTACATCTACCCGGGATCAAGTCGATCGGACGCGGACCCCA 632
Db 568 AAGTTGCTACTCAAGACCTACATCTACCCCGAATTAAGTCCGTCGCAACTGTTAAATCG 627
Qy 633 AAAGAGAGACTCATGTTTGGCGCAAT---CAAGCTGCCGACAGTTTGGCAAGTTGCC 689
Db 628 GTCCAGGAGCTTGTGTTTGGCTCCGTAAGCTAGCGCAGAGCATTAAGAGTATCCGT 687
Qy 690 ACTCCACTGGCAATCTTCGAGGAGTTTAT--AGCTGAGCGAGACCCACCTCTCCG-- 745
Db 688 CTGCTTTGAATGCTAGAGACTATGTCCAGTCTCGCAATAAGTCCCTTACCACGGAT 747
Qy 746 -----CCACTTTCTCTCATGCGATTTGGTCAAGCCGCTCC 779
Db 748 GACAGTCACAATACTCCOATTTATCTTACGCGCTTCTCTTTCGACCTGGTGGTCTTACC 807
Qy 780 GAGTCCCGAATCAAGGCTTACTGTATGGAACGCGCAGCTCGACCTGGCTCCCTCGAAGGT 839
Db 808 AAGTCTCGTGTCAAGATCTACCTCTGGAACGAATGTTCTGTTGCCAGCGATGGAAGAT 867
Qy 840 ATTTGGACTCTCAACGGCGACGGAACGATCCAGAGACACTGATGTTCTGGATGCGCTG 899
Db 868 CTTTGGACGCTTGGCGCGTTCGAGAAGATCAGTCCACTATTTAGGGGATTTGGAGATGATC 927
Qy 900 AGGAGCTGTGGAGCTATTTGCCGTCACGAGAGGTCTGTCTCAGTCCGCAACTGCTTT 959
Db 928 CGAGAACTTTGGGGTCTCTTAAACATGCTCTCTGTTGCGCGCTTACCTTGAGCCCTTAC 987
Qy 960 TAGGAGCGGTACTCTCACCGCAGGAGCTCCCTTCAATATAAATTTTACCTTTGTCT 1019
Db 988 TTGCCCTCGCGGCATTTCCCAATGAGCACTTCGTCCTATGGCCAATTCACCTTACAC 1047
Qy 1020 CTTAAAGCGCACTTCCGAACCAAGATCTATTTCCCTGCTTTTGGGCAAGACGACAAA 1079
Db 1048 CATATGATCGATACCAAGAACCGCAAGTGTACTTACTGTGTTCCGCAATGAATGATG 1107
Qy 1080 ACCATCGGGAAGGATTTGGCCACCTTCTTTGAGAGCAGAGGTTGGGTGGCTTGGCTAAG 1139
Db 1108 GAGGTGACTAATGCACTCACGAAATTTCTTCATGAGGCATGAATGAGCGATATGGCAAGT 1167
Qy 1140 AGCTATCCAGC 1150
Db 1168 AATATACAGC 1178

RESULT 14
AY259839
LOCUS Neotyphodium coenophialum 6055 bp DNA linear PLN 12-DEC-2003
DEFINITION Neotyphodium coenophialum L-tryptophan dimethylallyl transferase (dman) gene, dman-2 allele, complete cds.
ACCESSION AY259839
VERSION AY259839.1 GI:32402651
KEYWORDS
SOURCE Neotyphodium coenophialum

ORGANISM	Neotyphodium coenophialum
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Neotyphodium.
AUTHORS	1. (bases 1 to 6055)
TITLE	Wang, J., Machado, C. and Schardl, C.L.
JOURNAL	The determinant step in ergot alkaloid biosynthesis by a grass endophyte
REFERENCE	Fungal Genet. Biol. (2004) In press
AUTHORS	2. (bases 1 to 6055)
TITLE	Machado, C., Wang, J. and Schardl, C.L.
JOURNAL	Direct Submission
FEATURES	Submitted (17-MAR-2003) Plant Pathology, University of Kentucky, 201 Plant Sciences Bldg., Lexington, KY 40546-0312, USA
source	Location/Qualifiers
1. .6055	/organism="Neotyphodium coenophialum"
	/mol_type="genomic DNA"
	/isolate="e19; ATCC90664"
	/specific_host="Lolium arundinaceum (synonym: Pestuca arundinacea)"
gene	/db_xref="ATCC:90664"
	/db_xref="taxon:5047"
	<2961. .>4447
	/gene="dmaW"
	/allele="dmaW-2"
mRNA	join(<2961. .4132, 4199. .4320, 4389. .>4447)
	/gene="dmaW"
	/allele="dmaW-2"
CDS	/product="L-tryptophan dimethylallyl transferase"
	join(2961. .4132, 4199. .4320, 4389. .4447)
	/gene="dmaW"
	/allele="dmaW-2"
	/EC_number="2.5.1.34"
	/note="aromatic prenyltransferase;
	4-(gamma,gamma)-dimethylallyltryptophan synthase; DMAT synthase; 4-(3-methylbut-2-en-1-yl)-L-tryptophan synthase; dimethylallyl-pyrophosphate"
	/codon_start=1
	/product="L-tryptophan dimethylallyl transferase"
	/protein_id="AA881208.1"
	/db_xref="GI:32402652"
	/translation="MVLAKTLHQEVYQTLSETFDANRDLRWHSTAPMFQKILQTA NYGIYAQYQHLSTYKSHIIPFLGVPTGRGRLWLSILTRYGPFELSLNCSIVRYT YEPINATGSHLDPFNFTALWEALKLIDSGPDLQWFSYFKSLTLDANESTYLHS ONLVKEQIKTONKALDLKGDKEVLTYIYBELKSVATGSKVQELVFGSVKLAQKHK SIRPAPEMLEDYQSNKVPYTTDDSHNTEPLSRLSLCDLVSPFKSRVYLLEWVSL FAMEDLWTLGGRRDQSTIEGLEMIRELMGLNMGSLRAYPEPYLPGLAINEQLPS MANYTLHNDPIPEQVYTFVGMNDMEVTNATKFFMEHWSMASKYKALRESFP HHNYEALNVIHSYISFYSYRNKPYLSVLYLHSPETGEWPFVPEGLIAFDGCRDLTCYK"
ORIGIN	
Query Match	12.7%; Score 176.6; DB 8; Length 6055;
Best Local Similarity	50.7%; Pred. No. 5.3e-41;
Matches 583; Conservative	0; Mismatches 529; Indels 39; Gaps 5;
QY	33 GCAGCAACACTGCTGCGCAAGCCCTTCTAGTGTCTGAGTCAAGCCCTGAACTCTCGAAC 92
DB	2970 GCAAGCACTCCACGAGGAGTTTATCAACCCCTCAGCGAAACATTTGACTTTGCCAAC 3029
QY	93 AAGGACACACAAATATGGTGGTATAGACACAGCTCCGATGTTTGCCACCATGATGGGGG 152
DB	3030 AATGACCAAGGCTATGGTGGCACAGCAGCGGGCCAAATGTTCCAAAGATATCTCAAACT 3089
QY	153 GCCGGGTATGATGTTCCAGGCACAGTACAAGTTCCTCTGTATCCACCGTGAGTCAATC 212
DB	3090 GCTAACTATAGCATTTATGCTCAGTATCAACATCTGAGCATTTTATAAGCCATATCAIT 3149
QY	213 CCGCGGTGGTCCATACCCAGAAAGGTTCAGCCCATGCACTGGAAGAGTCAATCTCA 272
DB	3150 CTTTCTTTGGTGTCTATCTCTCAACAGAGTGG---CGAGCGGTGGCTAAGCATTTCTTACG 3206

QY	273 CGCTTCGAGCTTCCTTTGAGCTGAGCTTCAATTACTCAATACACTACTACGCTTTCGCA 332
DB	3207 AGATACGGAACCCCGTTTGGAGCTAAGTCTTAAATTCGCTGACTCCTAGTTTCGGAATACA 3266
QY	333 TTCGAGCCCTCGGTTCCCTGACGGGAAGGAAGATGATCCATTCAACACCCAGCGAATC 392
DB	3267 TACGAGCCCTATTAACGCGCGCAACTGGCAGCCATCTGGATCCGTTTCAACACTTTCGCTATC 3326
QY	393 AGGCTGTTCCTCCAGGACCTCAAGGCCATGTTCAGGGCTTGACCTGGGAATGTTGGAT 452
DB	3327 TGGGAGGCTCTAAAGAGCTTATAGATTTCCAGCAGCGCATAGACCTTCAATGTTTTC 3386
QY	453 CATTTCACTAAAGCATTTGGTTCGTTTCGGAGGAAGAGCTCGGACTCTGCTAGATFCGAGAT 512
DB	3387 TACTTTAAACAAGAGCTTACACTTTGAGGCAACGAGTCCAGTACCTGCACTCTCAAAAC 3446
QY	513 ATTGAGATCCCGCTTTCAGACACAGAACAACTGGCAGCCGATCTGGAGCCATCTGGC 572
DB	3447 TTGCTCAGGAACAGATCAAAACTCAAAACAGCTAGCTTAGACCTTAA---AGGTGAC 3503
QY	573 GATATTGCTTTGAAGACCTACATCTACCCGCGGATCAAGTCGATCGCGACCGGACCCCA 632
DB	3504 AAGTTGCTGACTCAAGACCTACATCTACCCGGAATTGAAGTCCGTCGCAACTGGTAAATCG 3563
QY	633 AAAGAGAGACTCATGTTTGACCGAAT---CAAGGCTCCGACAAAGTTTGGCAAGTTGCC 689
DB	3564 GTCCAGGAGCTTGTGTTTGGCTCCGTCGCAAGCTAGCGAGAAGCATAAAGATATCCGT 3623
QY	690 ACTCCACTGGCAATCTCCGAGGAGTTTAT---AGCTGAGCGAGCACCCACCTCTCTCGG-- 745
DB	3624 CTGCTTTGAAATGCTAGAGACTATGTCAGTCTCGCAATGAAGTCCCTACACCGAT 3683
QY	746 -----CCACTTCTCTCATGCGATTTGGTCAAGCGCTCC 779
DB	3684 GACAGTCACAATACTCCATTATCTTCACGCGCTTCTCTTTCGACCTGTGAGTCTTACC 3743
QY	780 GAGTCCCGGAATCAAGGCTTACTGTATGGAACGCCAGCTCGACCTGGCTCCATCGAAGGT 839
DB	3744 AAGTCTGCTCAAGATCTACTCTCTGGAAGATGCTCTGTTGCCAGCGATGGAAGAT 3803
QY	840 ATTTGGACTCTCAACGGGCGAAGCAAGATCCAGACACTCGGATGCTCTGATGGCGCTG 899
DB	3804 CTTTGGACGCTTGGCGCGCTCGAGAAGATCAGTCCACTATTGAGGATTTGGAGATGATC 3863
QY	900 AGGAGAGTGGCAGCTATTGCCGCTCAGGAGGCTGTGTCCACTGCCGCGAATGCTTT 959
DB	3864 CGAAGCTTTGGGGTCTCCTTAAATGCTCTCTGTTTGGCGCTTACCCCTGAGCCCTTAC 3923
QY	960 TACGAGCGGGTACTCTCACCGGAGGAGCTCCCTTCATTATAAATTTTACCTTGCTCT 1019
DB	3924 TTGCCCTCGCGGCAATCCCAATGAGCAACTCCGTCCTATGCGCAATTACACTTACAC 3983
QY	1020 CCTAAAAGCGCACTTCCGGAACACAGATCTATTTCCTGCTTTTGGGGAGAGACAAA 1079
DB	3984 CATAATGATCCGATACCAGAACCCGCAAGTGTACTTACTGTGTCGSCATGAATGATG 4043
QY	1080 ACCATCCGGAAGGATTTGGCCACCTTTTGAGAGCAGAGGTTGGGGTGGCTTGGCTAAG 1139
DB	4044 GAGTGACTAATGCACTCAGGAATTTCTTATGAGGATGAATGGAGCATATGGCAAGT 4103
QY	1140 AGCTATCCAGC 1150
DB	4104 AAATACAAAGC 4114
RESULT 15	
AY259837	AY259837 6640 bp DNA linear PLN 12-DEC-2003
LOCUS	Epichloe typhina x Neotyphodium lolii L-tryptophan dimethylallyl
DEFINITION	transferase (dmaW) gene, complete cds.
ACCESSION	AY259837
VERSION	AY259837.1 GI:32402647
KEYWORDS	

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Epichloe typhina x Neotyphodium lolii
Epichloe typhina x Neotyphodium lolii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocritales; Clavicipitaceae; Epichloe.
1 (bases 1 to 6640)
The determinant step in ergot alkaloid biosynthesis by a grass
endophyte
Wang, J., Machado, C. and Scharl, C.L.
Fungal Genet. Biol. (2004) In press
2 (bases 1 to 6640)
Wang, J., Machado, C. and Scharl, C.L.
Direct Submission
Submitted (17-MAR-2003) Plant Pathology, University of Kentucky,
201 Plant Sciences Bldg., Lexington, KY 40546-0312, USA
Location/Qualifiers
1 6640
/organism="Epichloe typhina x Neotyphodium lolii"
/mol_type="genomic DNA"
/isolate="Lpi"
/db_xref="taxon:71426"
<3111..>.4595
/gene="dnaW"
join(<3111..4282,4347..4468,4537..>.4595)
/gene="dnaW"
/product="L-tryptophan dimethylallyl transferase"
join(3111..4282,4347..4468,4537..4595)
/gene="dnaW"
/EC number="2.5.1.34"
/notes="aromatic prenyltransferase;
4-(gamma,gamma)-dimethylallyl-tryptophan synthase; DWAT
synthase; 4-(3-methylbut-2-enyl)-L-tryptophan synthase;
dimethylallyl-pyrophosphate"
/codon_start=1
/product="L-tryptophan dimethylallyl transferase"
/protein_id="AAP81206.1"
/db_xref="GI:32402648"
/translation="MVLAKTHQEVYTLSETFDPRANDORLWHSHPFQKILQTA
NYSIAQYQHSYKSHLPFLGVPTFRGSRWLSLLTRYGPFELSLNCSIVRYT
YEPINATGSHLPNTFAWEALKKLIDSPGIDQWFSYKQELTDLANSTYHS
QNLVQIKQIKLALDLKQDFVLKTYIYPELKSATQSKVOELVFGSRKLAQKH
SISPAFEMEDYVQSNKVPYTDSDHNTPLSRLLSCDLVSPTKSRVLYLERVSL
PAMEDLWLGREDOSTGLEIMELWGLNMSPLRAYPEYPLPLGAIPEQOLPS
MANYTHNDPIPEPOVYTFVGMNDMEVNTALTKFMRHWSDMASKVACLRSEFP
HNYEALNYLHSISYSYNNKPYLSVYLSHPETGWPVPEGLIAFDGCRDLICYK
"

Query Match 12.7%; Score 176.6; DB 8; Length 6640;
Best Local Similarity 50.7%; Pred. No. 5.3e-41;
Matches 583; Conservative 0; Mismatches 529; Indels 39; Gaps 5;
Qy 33 GCAGCAACCTGTCGCAAGCCCTTCTACGTGCTGAGTCAAGCCCTGACCTCTCGAAC 92
Db 3120 GCAAGACACTCCACCAGGAAGTTTATCAAAACCTCAGCGAAACATTTGACCTTGCCAAC 3179
Qy 93 AAGGACACACAAATGGTGTATAGCACAGCTCCGATGTTTGCACATGATGGGGGG 152
Db 3180 AATGACGAGGCTATGGTGGCAGACGCGGCCAATGTTCCAAAGATCTCCAAACT 3239
Qy 153 GCCGGCTATGATGTTACGACACAGTCAAGTTCCTGTGTATCCACCGTGAAGTCAATC 212
Db 3240 GCTAACTATAGCAATTTATGCTCAGTATCAACATCTGAGCATTTATATAAGCCATATCAIT 3299
Qy 213 CCGCGCTGGTCCATACCCAGAAAGGTCAGCCCATGCTGGAAGAGTCACTCACA 272
Db 3300 CCTTTCTTGGTGTCTATCCGACAGAAAGTGG---CGAGCGGTGGCTAAGCATCTTACG 3356
Qy 273 CGCTTCGAGCTCTCTTTTCGAGCTGAGCTTCAATTAATCTCCAAATCACTACTACGGTTTGA 332
Db 3357 AGATACGGAACCCGGTTTGAGCTAAGTCTTAATTGCTCTGACTCCATAGTTCGGTATACA 3416
Qy 333 TTCAGGCCCTCGGTTTCCTGACGCGGAACGAGAGATGATCCATTCAACACCCAGGCAATC 392

Db 3417 TACGAGCCTATTACGCCGCAACTGGAGCCATCTGGATCCGTTCAACACCTTGGCTATC 3476
Qy 393 AGGCCTGTTCTCCAGGACCTCAAGGCCATGTTCCAGGGCTTGACCTGGAATGTTCCGAT 452
Db 3477 TGGGAGGCTCTAAAAAGCTTATAGATCCAGCCAGCATAGACCTTCAATGTTTCC 3536
Qy 453 CATTTCACTAAGCAATTTGGTGTTCGAGAGAGAGCTCGGACTCTGCTAGATCGAGAT 512
Db 3537 TACTTTAAACAAGAGCTTACACTTTGACGCAAAACGAGTCCAGCTACCTCTCAAAAC 3596
Qy 513 ATTGAGATCCCGCTCTCAAGACACAGAAACAACTGGCAGCCGATCTGGAGCCATCTGGC 572
Db 3597 TTGTCAGGACAGATCAAAACTCAAAACAGGTAGCTAGACCTTAA---AGGTGAC 3653
Qy 573 GATATTGTCTTGAAGACTACATCTACCGCGGATCAAGTGGATCGGACGCGGACCCCA 632
Db 3654 AAGTTGCTGTCAGAGACTACATCTACCCCGAATGAAGTCCGTCGCAACTGGTAAATCG 3713
Qy 633 AAAGAGACTCATGTTTGGAGCAAT---CAAGGCTCCGACAAATTTGGCAAAAGTTGCC 689
Db 3714 GTCCAGGAGCTTGTGTTTGGCTCCGTCGCAAGCTAGCGCAAGATGAAGATTCOGT 3773
Qy 690 ACTCAGCTGGCAATCTCGAGGAGTTTAT--AGCTGAGCGAGCACCCACCTCTCTCG-- 745
Db 3774 CTTGCCCTTGAATGCTAGAGACTATGTCCAGTCTCGCAATAAAGTCCCTTACCACGAT 3833
Qy 746 -----CCACTTTCTCTCATCGGATTTGGTCAAGCCCTGCC 779
Db 3834 GACAGTCACAATACTCCATTAATCTTCACGCTTCTCTCTTCGACCTGGTGAGTCTTACC 3893
Qy 780 GAGTCCCGAATCAAGGTCTACTGTATGGAACGCCAGCTCCGACCTGGCTCCATCGAAGGT 839
Db 3894 AAGTCTGTCAGATCTACTCTCTGGAACGAATGGTCTCGTTGCCAGCGATGGAAGAT 3953
Qy 840 ATTGGAATCTCAACGGCGGACGGAACGATCCAGAGACACTGGATGGTCTGGATCGCGTG 899
Db 3954 CTTTGGAGCTTGGCGCGCTCGAGAGATCAGTCCACTATTGAGGGATTTGGAGATGATC 4013
Qy 900 AGGAGCTGTGGCAGCTATTGCCGCTCAGGAGGCTCTGTCCACTGCCGAACCTGCTTT 959
Db 4014 CGAGAACTTTGGGGTCTCCTTAAATGTCCTCTGTTTGGCGGCTACCTCGAGCCTTAC 4073
Qy 960 TACGAGCGGGTACCTCACCGCAGGAGAGCTCCCTTCAATTAATAATTTTACCTTGCT 1019
Db 4074 TTGCCCCCTCGGGCCCATTTCCCAATGAGCACTTCGCTCCATGGCCCAATTACACCTTACAC 4133
Qy 1020 CCTAAAGGCACTTCCGAAACCAGATCTATTTCCTGCTTTTGGSCAGAACACAA 1079
Db 4134 CATATGATGCGATACCAAGACCGCAAGTGTACTTGTGTCGGCATGAATGATATG 4193
Qy 1080 ACCATCGCGGAGGATTGGCCACCTTCTTTGAGAGCAGAGGTTGGGGTGGCTTGGCTAAG 1139
Db 4194 GAGGTGACTAATGCACTACGAAATCTTTCATGAGGCATGAATGGAGCATATGCAAGT 4253
Qy 1140 AGCTATCCAGC 1150
Db 4254 AAATACAAAGC 4264

Search completed: July 29, 2004, 00:10:31
Job time : 5592 secs

QM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 21:03:30 ; Search time 617 Seconds

(without alignments)
9591.151 Million cell updates/sec

Title: US-10-099-704-1

Perfect score: 1393
Sequence: 1 ccgaagctgcgaatggag.....cgaaaaaaaaaaaaaaaaaa 1393

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1393	100.0	1393	3	AA51712 A. oryzae
2	1052	75.5	1053	3	AAFL14653 Aspergill
3	582.4	41.8	600	7	ABZ53042 Aspergill
4	205.2	14.7	1347	6	ABK15520 DNA encod
5	205.2	14.7	1353	6	ABK15521 DNA encod
6	176.6	12.7	1598	6	ABK15523 DNA encod
7	175.6	12.6	1676	3	AAFL1858 Aspergill
8	173.4	12.4	1908	6	ABK15522 DNA encod
9	98.6	7.1	657	3	AAFL12651 Aspergill
10	43.6	3.1	2000	7	ADA71938 Rice gene
11	38.8	2.8	2000	7	ADA71938 Rice gene
12	35.6	2.6	1258	7	ACA31530 Prokaryot
13	34.8	2.5	484	5	ABV14861 Human pro
14	34.8	2.5	3237	4	AA506331 DNA encod
15	34.8	2.5	4254	7	ACA42624 Prokaryot
16	34.6	2.5	4380	5	ABAL4731 Human ner
17	34.4	2.5	1065	7	ABV72510 Nucleotid
18	34	2.4	14041	4	AAH48024 Internal
19	34	2.4	349980	6	ABQ81846 Bifidobac
20	33.8	2.4	4060	6	ABQ81281 Arabidops
21	33.8	2.4	58857	3	AA558471 Nucleotid
22	33.6	2.4	339	2	AAQ11993 Transform
23	33.6	2.4	1071	3	AA424403 Arabidops

24	33.6	2.4	1303	2	AAQ09317 Monkey tr
25	33.6	2.4	1559	2	AAQ20289 Sequence
26	33.6	2.4	1560	1	AA81084 Coding se
27	33.6	2.4	1560	2	AAQ03508 Simian Tr
28	33.6	2.4	1561	2	AAQ04908 Sequence
29	33.6	2.4	1561	2	AAQ03268 Simian tr
30	33.6	2.4	1561	2	AAQ29177 TGF-beta
31	33.6	2.4	1571	2	AAQ03269 Human tra
32	33.4	2.4	489	6	ABL99528 Target ca
33	33.4	2.4	511	8	ACH17350 Human adu
34	33.4	2.4	998	9	ADD35866 Human hep
35	33.4	2.4	1326	6	AA222696 Porcine t
36	33.4	2.4	11477	4	AAK68713 Human imm
37	33.4	2.4	11477	4	AAK71247 Human imm
38	33.4	2.4	11482	4	AAK68712 Human imm
39	33.4	2.4	11482	4	AAK71246 Human imm
40	33.4	2.4	96649	7	ACC45150 Human NAC
41	33.2	2.4	341	4	AAH72072 Human cer
42	33.2	2.4	350	4	AAH69197 Human cer
43	33.2	2.4	350	4	AAH71002 Human cer
44	33.2	2.4	3176	4	ABL05635 Drosophil
45	33.2	2.4	5415	4	ABL05634 Drosophil

ALIGNMENTS

RESULT 1

AA51712

ID AAA51712 standard; cDNA; 1393 BP.

XX AC AAA51712;

XX DT 31-OCT-2000 (first entry)

XX DE A. oryzae dimethylallyl-cycloacetyl-L-tryptophan synthase cDNA.

XX KW Dimethylallyl-cycloacetyl-L-tryptophan synthase; DCAT-S; beta-CPA;

XX KW toxin production; elimination; heterologous protein production;

XX KW filamentous fungi; ss.

XX OS Aspergillus oryzae.

XX FH Key Location/Qualifiers

FT CDS 15..1328

FT /*tag= a

FT /product= "dimethylallyl-cycloacetyl-L-

FT tryptophan synthase"

XX PN WO200039322-A1.

XX PD 06-JUL-2000.

XX PF 22-DEC-1999; 99WO-DK000726.

XX PR 23-DEC-1998; 98DK-00001726.

XX PR 27-MAY-1999; 99DK-00000745.

XX PA (NOVO) NOVO NORDISK AS.

XX PI Christensen BE, Mollgaard H, Kaasgaard S, Lehnbeck J;

XX DR WPI; 2000-452411/39.

XX DR P-PSDB; AAY96961.

XX PT Producing a polypeptide of interest such as a hormone or enzyme, comprising cultivating a mutant of a parent Aspergillus cell which produces less of at least one toxin of interest compared to the parent cell under the same conditions.

XX PS Claim 3i; Page 59-61; 66pp; English.

XX The A. oryzae dimethylallyl-cycloacetoacetyl-L-tryptophan synthase (DCAT-

CC S) is involved in the synthesis of beta-CPA, from cyclo-acetoacetyl-L-
 CC tryptophan and dimethylallylpyrophosphate, by its homology to a
 CC dimethylallyltryptophan synthase (DMAT-S) from *Claviceps purpurea*.
 CC *Aspergillus* host cells having a modification in the DCAT-S gene, leading
 CC to reduced or eliminated toxin production, are useful for expression of
 CC heterologous polypeptides of interest. Other toxins which may be reduced
 CC or eliminated comprise kojic acid, 3-nitropropionic acid, emodin, etc.
 CC The DCAT-S gene can be used to identify and disrupt similar genes in
 CC other filamentous fungal host strains such as *Trichoderma*, *Penicillium*
 CC and *Fusarium*
 XX
 SQ Sequence 1393 BP; 345 A; 387 C; 344 G; 317 T; 0 U; 0 Other;
 Query Match 100.0%; Score 1393; DB 3; Length 1393;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCAACATCTGTCGCAAGCCCTTCT 60
 DB 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCAACATCTGTCGCAAGCCCTTCT 60
 QY 61 ACCTGCTGAGTCAAGCCCTGAACCTCTCGAAACAAAGGACCACACAAATGGTATAGCA 120
 DB 61 ACCTGCTGAGTCAAGCCCTGAACCTCTCGAAACAAAGGACCACACAAATGGTATAGCA 120
 QY 121 CAGCTCGGATGTTGGCAACATGATGGGGGGCGGCTATGATGTTACGGACAGTACA 180
 DB 121 CAGCTCGGATGTTGGCAACATGATGGGGGGCGGCTATGATGTTACGGACAGTACA 180
 QY 181 AGTTCTCTGATACACCGTGAGTGCATCATCCCGGCGTGGGTCCATACCCAGAAAGG 240
 DB 181 AGTTCTCTGATACACCGTGAGTGCATCATCCCGGCGTGGGTCCATACCCAGAAAGG 240
 QY 241 GTGAGCCCATGCACTGGAAGAGTCACTCACAGCTTCGGAATTCCTTCGAGCTGAGCT 300
 DB 241 GTGAGCCCATGCACTGGAAGAGTCACTCACAGCTTCGGAATTCCTTCGAGCTGAGCT 300
 QY 301 TCAATTACTCCAAATCACTACTACGTTGCGATTCGAGCCCTCGGTCCTCGACGGGAA 360
 DB 301 TCAATTACTCCAAATCACTACTACGTTGCGATTCGAGCCCTCGGTCCTCGACGGGAA 360
 QY 361 CGAAGGATGATCCATTCACACACCCAGGCAATCAGGCTCTGTTCTCCAGGACCTCAAGGCCA 420
 DB 361 CGAAGGATGATCCATTCACACACCCAGGCAATCAGGCTCTGTTCTCCAGGACCTCAAGGCCA 420
 QY 421 TGGTTCAGGGCTTGACCTGGATGTTGCGATCAATTTCACTAAGCATTTGCTTCGG 480
 DB 421 TGGTTCAGGGCTTGACCTGGATGTTGCGATCAATTTCACTAAGCATTTGCTTCGG 480
 QY 481 AGAAGAGGCTCGGACTCTGCTAGATCGAGATATTCAGATCCCGCTTCAGAGACACAGA 540
 DB 481 AGAAGAGGCTCGGACTCTGCTAGATCGAGATATTCAGATCCCGCTTCAGAGACACAGA 540
 QY 541 ACAAATCTGGCAGCGATCTGGAGCCATCTGGCGATATTTGTTGAAGACCTACATCTACC 600
 DB 541 ACAAATCTGGCAGCGATCTGGAGCCATCTGGCGATATTTGTTGAAGACCTACATCTACC 600
 QY 601 CGCGGATCAAGTCTGATCGGACCGGACCCCAAGAGAGAGATCTATGTTTGAAGCAATCA 660
 DB 601 CGCGGATCAAGTCTGATCGGACCGGACCCCAAGAGAGAGATCTATGTTTGAAGCAATCA 660
 QY 661 AGGCTCCGCAAGATTTGGCAAGTTGGCCATCTCCACTGGCAATTCCTCGAGGAGTTTATAG 720
 DB 661 AGGCTCCGCAAGATTTGGCAAGTTGGCCATCTCCACTGGCAATTCCTCGAGGAGTTTATAG 720
 QY 721 CTGAGCGGACCCACCCCTCTCGGCACTTTCTCTCATGCGGATTTGGTCAAGCGCTCG 780
 DB 721 CTGAGCGGACCCACCCCTCTCGGCACTTTCTCTCATGCGGATTTGGTCAAGCGCTCG 780
 QY 781 AGTCCCGAATCAAGGCTCTACTGTATGGAACCGGACGCTCGACCTGGCTCCATCGAAGGTA 840
 DB 781 AGTCCCGAATCAAGGCTCTACTGTATGGAACCGGACGCTCGACCTGGCTCCATCGAAGGTA 840

QY 841 TTTGGACTCTCAACGGGCGACGGAACGATCCAGAGACACTGGATGGTCTGGATGGCGCTGA 900
 DB 841 TTTGGACTCTCAACGGGCGACGGAACGATCCAGAGACACTGGATGGTCTGGATGGCGCTGA 900
 QY 901 GGGAGCTCTGGAGCTATTCCCGCTCAGCGAGGCTGTGTCCACTGCGCAACTGCTTTT 960
 DB 901 GGGAGCTCTGGAGCTATTCCCGCTCAGCGAGGCTGTGTCCACTGCGCAACTGCTTTT 960
 QY 961 ACAGCCGGGTACCTCACCAGCAGGAGCAGCTCCCTTCATTATAAAATTTTACCTTGTCTC 1020
 DB 961 ACAGCCGGGTACCTCACCAGCAGGAGCAGCTCCCTTCATTATAAAATTTTACCTTGTCTC 1020
 QY 1021 CTAAGAGCGGACTTCCGAAACCAAGATCTATTCCCTGCTTTTGGGCAACGACAAA 1080
 DB 1021 CTAAGAGCGGACTTCCGAAACCAAGATCTATTCCCTGCTTTTGGGCAACGACAAA 1080
 QY 1081 CCATCGCGGAAGGATTTGGCCACCTCTTTGAGAGCAGAGGTTGGGCTTGGCTTAAGA 1140
 DB 1081 CCATCGCGGAAGGATTTGGCCACCTCTTTGAGAGCAGAGGTTGGGCTTGGCTTAAGA 1140
 QY 1141 GCTATCAGCGGATTTGGCATCTCTTCTTACAAAGGGGAAAAACCGTACATGAGTGTGACTCC 1200
 DB 1141 GCTATCAGCGGATTTGGCATCTCTTCTTACAAAGGGGAAAAACCGTACATGAGTGTGACTCC 1200
 QY 1201 TGCAGGCGTGTGATCTCTCTTCTTACAAAGGGGAAAAACCGTACATGAGTGTGACTCC 1260
 DB 1201 TGCAGGCGTGTGATCTCTCTTCTTACAAAGGGGAAAAACCGTACATGAGTGTGACTCC 1260
 QY 1261 ATACCTTCGAAGCGTTTCAGTGTCTGCTGCCAGAGGTTGGCTATGTCAGATGGCCACA 1320
 DB 1261 ATACCTTCGAAGCGTTTCAGTGTCTGCTGCCAGAGGTTGGCTATGTCAGATGGCCACA 1320
 QY 1321 ATCCTTAGGACTAGTTTATCCCTTCATTCTATGCATCCGTTGAAATGTTGGTGGCAAAA 1380
 DB 1321 ATCCTTAGGACTAGTTTATCCCTTCATTCTATGCATCCGTTGAAATGTTGGTGGCAAAA 1380
 QY 1381 AAAAAAAAAAAAAA 1393
 DB 1381 AAAAAAAAAAAAAA 1393
 RESULT 2
 AAF14653
 ID AAF14653 standard; cDNA; 1053 BP.
 XX
 AC AAF14653;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE *Aspergillus oryzae* EST SEQ ID NO:7176.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;
 KW *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS *Aspergillus oryzae*.
 XX
 PN WO200056762-A2.
 XX
 XX 28-SEP-2000.
 PD
 XX 22-MAR-2000; 2000WO-US007781.
 PF
 XX 22-MAR-1999; 99US-00273623.
 PR
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 PI Berka RM, Ray MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX

XX PT Monitoring differential expression of genes in filamentous fungal cells
PT PT uses fluorescence-labeled nucleic acids isolated from the cells and a
XX PT substrate of expressed sequence tags.
XX
XX Claim 88; Page 2910-2911; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring the
CC global expression of genes from FF cells allows the production potential
CC of the microorganisms to be improved. New genes may be discovered,
CC possible functions of unknown open reading frames can be identified and
CC gene copy number variation and stability can be monitored. The expression
CC of genes can be used to study how FF cells adapt to changes in culture
CC conditions, environmental stress, spore morphogenesis, recombination,
CC metabolic or catabolic pathway engineering. Using ESTs provides several
CC advantages over genomic or random cDNA clones including elimination of
CC redundancy as one spot on an array equals one gene or open reading frame,
CC and organisation of the microarrays based on function of the gene
CC products to facilitate analysis of the results. AAF07478 to AAF11247
CC represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents
CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from
CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from
CC *Trichoderma reesei*, which are all specifically claimed in the present
CC invention
XX
XX SQ Sequence 1053 BP; 256 A; 305 C; 259 G; 232 T; 0 U; 1 Other;

XX Query Match 75.5%; Score 1052; DB 3; Length 1053;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1052; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCACTGCTGCCAAGCCCTTCT 60
DB 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCACTGCTGCCAAGCCCTTCT 60

QY 61 ACCTGCTGAGTCAAGCCCTGAACTCTCGAAGCAGCACTGCTGCCAAGCCCTTCT 120
DB 61 ACCTGCTGAGTCAAGCCCTGAACTCTCGAAGCAGCACTGCTGCCAAGCCCTTCT 120

QY 121 CAGCTCCGATGTTGCCACATGATGCGGGGCGCGCTATGATGTTACGCACAGTACA 180
DB 121 CAGCTCCGATGTTGCCACATGATGCGGGGCGCGCTATGATGTTACGCACAGTACA 180

QY 181 AGTTCTCTGTATCCACCGTGAGGTATCATCCCGCGTGGTCCATACCCAGAAAGG 240
DB 181 AGTTCTCTGTATCCACCGTGAGGTATCATCCCGCGTGGTCCATACCCAGAAAGG 240

QY 241 GTCAGCCCATGCACTGGAAGAGTCACTCTCACAGCTTCGGACTTCCTTCGAGCTGAGCT 300
DB 241 GTCAGCCCATGCACTGGAAGAGTCACTCTCACAGCTTCGGACTTCCTTCGAGCTGAGCT 300

QY 301 TCAATTACTCCAAATCACTACTACGTTTGGATTCGAGCCCTCGTTCCTTCGAGCGGNA 360
DB 301 TCAATTACTCCAAATCACTACTACGTTTGGATTCGAGCCCTCGTTCCTTCGAGCGGNA 360

QY 361 CGAAGGATGATCCATTCAACACCCAGGCAATCAGGCTGTTCTCCAGGACCTCAAGGCCA 420
DB 361 CGAAGGATGATCCATTCAACACCCAGGCAATCAGGCTGTTCTCCAGGACCTCAAGGCCA 420

QY 421 TGGTTCCAGGGCTTGAACCTGGAATGTTTCGATCTTCACTAAAGCATTTGCTGTTTCGG 480
DB 421 TGGTTCCAGGGCTTGAACCTGGAATGTTTCGATCTTCACTAAAGCATTTGCTGTTTCGG 480

QY 481 AGGAAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGCTTCAAGACACAGA 540
DB 481 AGGAAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGCTTCAAGACACAGA 540

QY 541 ACAAACTGGCAGCGATCTGGAGCCATCTGGCGATATTGCTTGAAGACCTACATCTACC 600
DB 541 ACAAACTGGCAGCGATCTGGAGCCATCTGGCGATATTGCTTGAAGACCTACATCTACC 600

QY 601 CGCGGATCAAGTGCATCGCGACCGGGACCCCAAAAGAGAGACTCATCTTTGACGCAATCA 660
DB 601 CGCGGATCAAGTGCATCGCGACCGGGACCCCAAAAGAGAGACTCATCTTTGACGCAATCA 660

QY 661 AGGCTGCGGCAAGTTTGGCAAGTTGCGCACTCCACTGCGCAATCTCGAGGAGTTTATAG 720
DB 661 AGGCTGCGGCAAGTTTGGCAAGTTGCGCACTCCACTGCGCAATCTCGAGGAGTTTATAG 720

QY 721 CTGAGCGAGCAGCACCACCTCTCTCGGCCACTTTCTCTCATGCGAATTTGGTCAAGCCGTCG 780
DB 721 CTGAGCGAGCAGCACCACCTCTCTCGGCCACTTTCTCTCATGCGAATTTGGTCAAGCCGTCG 780

QY 781 AGTCCCAATCAAGTCTACTGTATGAAAGCCAGCTCGACCTGGCTCCATCGAAGGTA 840
DB 781 AGTCCCAATCAAGTCTACTGTATGAAAGCCAGCTCGACCTGGCTCCATCGAAGGTA 840

QY 841 TTTTGACTCTCAACGGGCGACGGAACGATCCAGAGACACTGGATGGTCTGGATCGGCTGA 900
DB 841 TTTTGACTCTCAACGGGCGACGGAACGATCCAGAGACACTGGATGGTCTGGATCGGCTGA 900

QY 901 GGGAGCTGTGGCAGCTATTGCGCGCTGACGGAGGCTCTGTCTCCACTGCCGACTGCTTTT 960
DB 901 GGGAGCTGTGGCAGCTATTGCGCGCTGACGGAGGCTCTGTCTCCACTGCCGACTGCTTTT 960

QY 961 ACGAGCGGGTACCTCACCAGGAGCAGCTCCCTTCAATTAATAATTTTACCTTGTCTC 1020
DB 961 ACGAGCGGGTACCTCACCAGGAGCAGCTCCCTTCAATTAATAATTTTACCTTGTCTC 1020

QY 1021 CTAAAGCGCACTTCCCGAACCCAGATCTATT 1053
DB 1021 CTAAAGCGCACTTCCCGAACCCAGATCTATT 1053

XX RESULT 3
XX ABZ53042
XX ID ABZ53042 standard; cDNA; 600 BP.
XX AC ABZ53042;
XX DT 28-MAR-2003 (first entry)
XX DE *Aspergillus oryzae* polynucleotide SEQ ID NO 2155.
XX KW *Aspergillus oryzae*; fermentation; fungus; industrial; EST;
XX KW expressed sequence tag; gene; ss.
XX OS *Aspergillus oryzae*.
XX PN WO200279476-A1.
XX PD 10-OCT-2002.
XX PF 22-MAR-2002; 2002WO-IB000890.
XX PR 30-MAR-2001; 2001JP-00098371.
XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX PA (NARE-) NAT RES INST BREWING.
XX PA (NORQ) NAT FOOD RES INST MIN AGRIC.
XX PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
XX PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX WPI; 2003-046817/04.
XX
XX PT Detection of expression of specific *Aspergillus* genes for monitoring the
XX fermentation and growth conditions of the fungus, using DNA probes.

PS Claim 1; SEQ ID NO 2155; 48pp + Sequence Listing; Japanese.
 XX The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridizing
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus,
 CC especially of *Aspergillus oryzae* which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 600 BP; 144 A; 176 C; 146 G; 134 T; 0 U; 0 Other;

Query Match 41.8%; Score 582.4; DB 7; Length 600;
 Best Local Similarity 99.8%; Pred. No. 2.9e-175;
 Matches 583; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 286 CTTTCGAGCTGAGCTCAATTACTTCAATCACTACTAGCTTGCATTTCGAGCCCTCG 345
 DB 17 CTTTCGAGCTGAGCTTCAATTACTTCAATCACTACTAGCTTGCATTTCGAGCCCTCG 76
 QY 346 GTTCCCTGACGGGAACGAGGATGATCCATTCAACACCCAGGCAATTCAGGCTGTTCCTCC 405
 DB 77 GTTCCCTGACGGGACGAGGATGATCCATTCAACACCCAGGCAATTCAGGCTGTTCCTCC 136
 QY 406 AGACCTCAAGGCGATGTTCCAGGCTGACCTCGAATGTTTCGATCACTTCACTAAG 465
 DB 137 AGACCTCAAGGCGATGTTCCAGGCTGACCTCGAATGTTTCGATCACTTCACTAAG 196
 QY 466 CATTGCTGTTTCGGAGGAGGCTCGGACTCTGTAGATCGAGATATTGATGCCCG 525
 DB 197 CATTGCTGTTTCGGAGGAGGCTCGGACTCTGTAGATCGAGATATTGATGCCCG 256
 QY 526 TCTTCAAGACACAGAACAACTGGCAGCCGATTCGAGCCATTCGCGATATTGTCTGA 585
 DB 257 TCTTCAAGACACAGAACAACTGGCAGCCGATTCGAGCCATTCGCGATATTGTCTGA 316
 QY 586 AGACCTACATCTACCGCGGATCAAGTCGATCGGACCGGACCCCAAGAGAGACTCA 645
 DB 317 AGACCTACATCTACCGCGGATCAAGTCGATCGGACCGGACCCCAAGAGAGACTCA 376
 QY 646 TGTTCGACGCAATCAAGGCTGCGACAAAGTTGGCAAGTTGCACTCCGCAATCC 705
 DB 377 TGTTCGACGCAATCAAGGCTGCGACAAAGTTGGCAAGTTGCACTCCGCAATCC 436
 QY 706 TCAGAGATTATAGTCGAGGAGCACCACCCCTCTCGGCACTTCTCATCGATT 765
 DB 437 TCAGAGATTATAGTCGAGGAGCACCACCCCTCTCGGCACTTCTCATCGATT 496
 QY 766 TGGTCAAGCGCTCGAGTCCCGAATCAAGTCTACTGTATGGAACGCCAGCTCGACCTGG 825
 DB 497 TGGTCAAGCGCTCGAGTCCCGAATCAAGTCTACTGTATGGAACGCCAGCTCGACCTGG 556
 QY 826 CTTCCATCGAAGTATTGGAATCTCTCAACGGGCGAGCAACGAT 869
 DB 557 CTTCCATCGAAGTATTGGAATCTCTCAACGGGCGAGCAACCAT 600

RESULT 4
 ABK15520
 ID ABK15520 standard; DNA; 1347 BP.
 XX
 AC ABK15520;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE DNA encoding dimethylallyltryptophan synthase (DmaW) version #1.
 XX

KW endophyte; forage; pasture; turf; land reclamation; soil conservation;
 KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
 KW Balansia; Balansiolepis; Echinodopsis; Atkinsonella; Myriogenospora;
 XX Neotyphodium; Parepichloe; knockout; antisense technology; gene; ds.
 OS Neotyphodium coenophialum.
 FH Key Location/Qualifiers
 FT CDS 1..1347
 FT /*tag= a
 FT /product= "dmaW"
 XX /note= "Dimethylallyltryptophan synthase"
 PN US6335188-B1.
 XX
 PD 01-JAN-2002.
 XX
 PF 03-MAR-2000; 2000US-00518657.
 XX
 PR 22-MAR-1999; 99US-0125490P.
 XX
 PA (KENT) UNIV KENTUCKY RES FOUND.
 XX
 PI Schardl CL, Wang J;
 XX WPI; 2002-163205/21.
 DR P-PSDB; AAU76412.
 XX
 PT New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 PT molecule) from fungi that are symbionts of commercially important
 PT grasses, useful to engineer ergot alkaloid-deficient symbionts.
 XX
 PS Claim 1; Col 9-12; 16pp; English.
 CC The invention describes an isolated dimethylallyltryptophan synthase
 CC (dmaW) nucleic acid (I). (I) is a dmaW sequence from epiphyte fungi
 CC Neotyphodium coenophialum symbiont of commercially significant grasses
 CC and is useful for expressing dmaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a dmaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epichloe that lack dmaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of dmaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansia, Balansiolepis,
 CC Echinodopsis, Atkinsonella, Myriogenospora, Neotyphodium, and
 CC Parepichloe, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This sequence encodes a
 CC dimethylallyltryptophan synthase (dmaW) described in the method of the
 CC invention
 XX
 SQ Sequence 1347 BP; 363 A; 345 C; 288 G; 351 T; 0 U; 0 Other;

Query Match 14.7%; Score 205.2; DB 6; Length 1347;
 Best Local Similarity 50.8%; Pred. No. 2.4e-54;
 Matches 647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;
 QY 33 GCAGCAACACTGCTGCCAAGCCCTTCTAGCTGCTAGTCAGCCCTGACCTCTCGAAC 92
 DB 10 GCAAGACACTCCACAGGAGTATATACCTTAGCGAACAATTTGACTTTGCCAAC 69
 QY 93 AAGGACCAACAAAAATGGTGGTATAGCAGAGCTCCGATGTTTCCACCATGATGGCGGG 152
 DB 70 AATGACCAGAGGCTATGTTGSCACAGCAGCGGCCAATGTTGAAAAGATGCTCCAACT 129
 QY 153 CGCGGCTATGATTTACGACAGTACAGTTCCTCTGTATCCACCGTAGGTCATCATC 212

130 GCTAATATAGCATTTGATGCTCAGTATCGACATCTGGCACTTTATAAGAGCCATGTCATT 189
213 CCGGCGTTGGTCCATATACCGAAGAAAGGTCAGCCATGCACTGCAAGAGTCATCTCACA 272
190 CTTTTCCTTGGTCTATCTCAAGAAAGTGG---CGAGCGGTGGCTAAGCATTTCTTAGC 246
273 CGCTTCGGACTTCCCTTTGAGCTGAGCTTCAATTAATCTCCAAATCACTACTACGGTTTGA 332
247 AGATACGGAAACCCGTTTGGCTAGTCTTAATGCTCGGACTCGTAGTTCGGTATACA 306
333 TTGAGCCGCTCGTTTCCCTGACGGGAACGAAGGATGATCCATTTCAACACCCAGGCAATC 392
307 TAGGAGCCTATTAAACCCGCACTGGCAGTCTATCTGGATCCGTTTAAACACTTTTCGCTATC 366
393 AGGCTGTCTCCAGAGCCTCAAGGCCATGGTTCCAGGCTTTCAGGCTTTCAGCTGAATGGTTCGAT 452
367 TGGAGGCCCTGAAAAGCATATTGAGTCCGAGCCAGGATAGACTTGAATGGTTTCT 426
453 CATTTCACTAAAGCATTGGTTCGTTCCGAGGAAGAGCTCGGACTCTGCTAGATCGAGAT 512
427 TACTTTAAACAAGAGCTTACACTTTGACGCAACGAGTCCACTGCTGCACTCGCAAAAC 486
513 AITGAGATCCCCGCTCTCAAGACAGACAACTGGCAGCCGATCTGGAGCCATCTGGC 572
487 TTGGTTAAGAAACAGATCAAACTCAAAACAAGCTCGCTTTGGACCTTAA---AGTGAC 543
573 GATATTGTCTTGAAGACCTACATCTACCCGCGATCAAGTCGATCGCGACCGGACCCCA 632
544 AAGTTCGTACTGAAGACCTACATCTACCCGGAATTTGAAGTCCGTCGCACTGCTAAATCG 603
633 AAGAGAGACTGATGTTTGAAGCAAT---CAAGGCTGCGCAAGATTTGGCAAGTTGCC 689
604 GTCCAGGAGCTCGTGTGTTGGCTCCGTCGCAAGCTAGCGCAGAAAGAGATATCCGT 663
690 ACTCCACTGCAATCTCTGAGGAGTTTAT--AGCTGAGCGAGCACCCACCTCTCGG-- 745
664 CTTGCTTTGAAATGCTAGAAGACTATGTCAGTCTCGCAATAAATCTCTACCGAGAT 723
746 -----CCACTTCTCTCATGCGATTTGCTCAAGCGCTC 779
724 GACAGTCACAATACTCTATTATCTTACGCGCTTCTCTTGGACCTGATAAGTCTTACC 783
780 GAGTCCGGAATCAAGCTCTACTGTATGGAAGCCAGCTCGACTGCTGGCTCCATCGAGGT 839
784 AAGTCTCGTCAAGATCTACTCTCTGGAACGAATGGTCTCGTTGCCAGCGATGAAGAT 843
840 ATTTGGACTCTCAACGGCGCAGCAACGATCCAGAGACACTGGAATGCTTGGATGGCTG 899
844 CTTTGGACGCTTCGCGCGCTCGAGAAGATCACTCCACTATTGAGGGATTTGAGATGATC 903
900 AGGAGCTGTGGAGCTATTGCCCGTCAAGGAGGTCTGTGTCCTCACTGCGCACTCTT 959
904 CGAGAACTTTGGGCTCTCCTCAACATGTCTCTCTGTTGGCGCTTACCCCTGAGCCTTAC 963
960 TAGAGCGGGTACCTCACCGCAGGAGCAGCTCCCTTCAATATAATTTTACCTTGTCT 1019
964 TTGCCCCCTGGCGCCATTCCTATGAGCACTTCCGTCATGCGCAATTACACCTTACAC 1023
1020 CCTAAAGCCGATTTCCCGAACACAGATCTATTTCCTGCTTTTGGGAGAACGCAAAA 1079
1024 CATAATGATCCCATACAGAACCGCAAGTACTTACTTGTGTTGGCATGAATGATG 1083
1080 ACCATCGGAGAGATTTGGCCACTCTTTTGAGCAGAGGTTGGGTCGCTGGCTAAG 1139
1084 GAGGTGATATGCACTCAAGCATTTCTCATGAGGATGAATGGAGCGATATGGCAATG 1143
1140 AGCTATCCAGCGGATTTGGCATCTACTATCCCGATGTGGACCTGCGAGCCGCAATCAC 1199
1144 AAATACAAAGCTTGCCTCAGGGAATCTTTCCGCGATCATGATTACGAGCCCTGATTTAT 1203
1200 CTGAGCGGTGATCTCTCTTTTACAGGGGAAAAACCGTACATGATGTTGACCTC 1259
1204 ATCCACTCGTACATTTCTCTCTCCTACCGAAAGAACAGCCATATTAAAGTGTGATCTC 1263

QY 1260 CATACCTTCGAAGC 1273
DB 1264 CACTCCITGAAAC 1277
RESULT 5
ABK15521
ID ABK15521 standard; DNA; 1353 BP.
XX ABK15521;
AC ABK15521;
XX 08-MAY-2002 (first entry)
XX DNA encoding dimethylallyltryptophan synthase (DmaW) version #2.
DE Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
KW endophyte; forage; pasture; turf; land reclamation; soil conservation;
KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
KW Balansia; Balansiolepis; Echinodopsis; Atkinsonella; Myriogenospora;
KW Neotyphodium; Parepichloe; knockout; antisense technology; gene; ds.
XX Neotyphodium coenophialum.
XX
XX Key Location/Qualifiers
FH CDS 1..1353
FT /*tag= a
FT /product= "dmaW"
FT /note= "Dimethylallyltryptophan synthase"
XX
PN US6335188-B1.
XX
XX 01-JAN-2002.
XX 03-MAR-2000; 2000US-00518657.
XX 22-MAR-1999; 99US-0125490P.
XX (KENT) UNIV KENTUCKY RES FOUND.
XX Schardl CL, Wang J;
XX WPI; 2002-163205/21.
XX P-PSDB; AAU76413.
XX New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW molecule) from fungi that are symbionts of commercially important grasses; useful to engineer ergot alkaloid-deficient symbionts.
XX
XX Claim 1; Col 13-16; 16pp; English.
XX The invention describes an isolated dimethylallyltryptophan synthase (dmaW) nucleic acid (I). (I) is a dmaW sequence from epiphyte fungi Neotyphodium coenophialum symbiont of commercially significant grasses and is useful for expressing dmaW in a cell by recombinant techniques. (I) is also useful for: identifying endophytes e.g. from commercial plants used in forage, pasture, turf, land reclamation and soil conservation that contain or lack a dmaW gene and producing increased amount of ergot alkaloids involving expressing (I) in a host fungal cell so that the copy number of mRNA derived from transcription of the nucleic acid molecule is increased, allowing the host fungal cell to grow under appropriate growth conditions, thus causing increased production of ergot alkaloid. Probes derived from (I) are useful for identifying Neotyphodium or Acremonium or Epichloe that lack dmaW and therefore are unlikely to produce ergot alkaloids. The primers are useful for amplifying segments of dmaW from fungi in family Clavicipitaceae. (I) is also useful for identifying related sequences such as from Balansia, Balansiolepis, Echinodopsis, Atkinsonella, Myriogenospora, Neotyphodium, and Parepichloe, or natural or induced mutants. A knockout construct of (I) or antisense construct is useful for engineering ergot alkaloid-deficient fungal symbionts (endophytes of plants). This sequence encodes a dimethylallyltryptophan synthase (dmaW) described in the method of the invention

XX SQ Sequence 1353 BP; 373 A; 345 C; 279 G; 356 T; 0 U; 0 Other;

Query Match 14.7%; Score 205.2; DB 6; Length 1353;
 Best Local Similarity 50.8%; Pred. No. 2.4e-54;
 Matches 647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;

33 GCACACACTGCTGCCAAGCCCTTCTAGCTGCTGAGTCAAGCCCTGACCTCTCGAAC 92
 |||||
 10 GCACACACTGCTGCCAAGCCCTTCTAGCTGCTGAGTCAAGCCCTGACCTCTCGAAC 69
 |||||
 93 AAGGACACACAAATGGTGTATAGCACAGCTCCGATGTTGCCACCATGATGGCGGG 152
 |||||
 70 AATGACACAGAGCTATGGTGGCAGCAGCGGCCCAATGTTCCAAAGATACCTCCAACT 129
 |||||
 153 GCCGGCTATGATGTTACGACACAGTACAGTCTCTGTATCCACCGTGAGGTCAATC 212
 |||||
 130 GCTAACTATAGCATTTATGCTCAGTATCAACATCTGAGCATTTATAAAGCCATATCAT 189
 |||||
 213 CCAGCGTGTGGGTCCATACCCAGAAAGGCTCAGCCATGCACTGGAAGAGTCACTCCACA 272
 |||||
 190 CTTTCTTCTGCTGCTATCTACAGAGTGG---CGAGCGGTGGCTAAGCATCTCTAGC 246
 |||||
 273 CGCTTCGAGCTTCGTTTCGAGCTGAGCTTCAATTAATCACTCAATCACTACGTTTGA 332
 |||||
 247 AGATACGGAACCCCGTTGAGCTTAAGCTTTAATTTGCTGACTCCATAGTTCGGTATAC 306
 |||||
 333 TTCGAGCCCTCGGTTCCCTGACGGGAAACGAGGATGATCCATTCAACACCCAGGCAATC 392
 |||||
 307 TACGAGCTTAATACGCGGCACTGSCAGCCATCTGGATCCGTTCAACACTTTCGTATC 366
 |||||
 393 AGGCTGTTTCCAGGACCTCAAGGCCATGTTTCCAGGGCTTGAACCTGGAATGGTTTCAT 452
 |||||
 367 TGGAGGCTCTAAAAAGCTTATAGATTTCCAGCCAGGATAGACCTTCAATGGTTTTC 426
 |||||
 453 CATTTCATCAAGCATTTGTCGTTTCGAGGAGAGCTCGGACTCTGCTAGATCGAGT 512
 |||||
 427 TACTTTAACAAGAGCTTACACTTGACGCAACAGTCCAGTACCTGCACTCTCAAAAC 486
 |||||
 513 ATTGAGATCCCGTCTTCAAGACACAGAACAACTGGCAGCCGATCTGGAGCCATCTGGC 572
 |||||
 487 TTGGTCAAGGAACAGATCAAACTCAAAACAGCTAGCTAGCTTAA---AGTGAAC 543
 |||||
 573 GATATTGCTTGAAGACCTACATCTACCGCGGATCAAGTCGATCCGACCGGACCCCA 632
 |||||
 544 AGTTCGTACTCAAGACCTACATCTACCCGGAATGAAGTCCGTCCCAACTGGTAATCG 603
 |||||
 633 AAAGAGAGACTCATGTTTGACGCAAT---CAAGGCTGCGGACAAAGTTTGGCAAGTTGCC 689
 |||||
 604 GTCCAGGAGCTTGTGTTTGGCTCGCTCCGAGCTAGCCAGCAGNAGCATAAAGATATCCGT 663
 |||||
 690 ACTCCACTGCAATCTTCGAGGAGTTTAT--AGCTGAGGAGACCCACCTCTCGG-- 745
 |||||
 664 CTTGCTCTTGAATGTAGAGACTATGTCAGTCTCCGCAATAAAGTCCCTACCAAGGAT 723
 |||||
 746 -----CCACTTCTCTCATGCGATTGGTCAAGCCGCTCC 779
 |||||
 724 GACAGTCACATACTCCATATCTTCAAGCTTCTCTTGGACCTGTGAGTCTTACC 783
 |||||
 780 GAGTCCGGAATCAAGTCTACTGTATGGAACGCCAGCTCGACCTGGCTCCATCGAAGT 839
 |||||
 784 AAGTCTGCTGCAAGATCTACTCTCTGGAACGAATGCTCTGTTGCGAGCATGGAAGT 843
 |||||
 840 ATTGAGCTCTCAACGGGCGGACGAGATCCAGACACACTGGATGGTCTGATCGGCTG 899
 |||||
 844 CTTTGGACGCTTGGCGCGCGTCCAGAGATCACTCACTTAATGAGGATGAGATGATC 903
 |||||
 900 AGGAGCTGTGGCAGTATTTGCCGCTCAGCGAGGCTGTGTCCACTGCGCAACTGCTTT 959
 |||||
 904 CGAGAACTTTGGGCTCTCTTAACATGTCTCTGGTTTGGCGGCTTACCCTGAGCCTTAC 963
 |||||
 960 TACGAGCGGCTACCTCAGCGAGGAGCTCCCTTCAATTAATTAATTTACCTTGTCT 1019
 |||||

Db 964 TTGCCCTCGCGCCATTCCCAATGAGCAACTTCCGTCCATGGCAATTACACCTTACAC 1023
 QY 1020 CCTAAAGCGCACTTCCCGAACCAAGATCTATTTCCCTGCTTTTGGGCGAGACGACAAA 1079
 |||||
 Db 1024 CATATGATCGGATACCAAGAACCGCAAGTGTACTTTACTGTGTTCCGCGCATGAATGATG 1083
 |||||
 QY 1080 ACCATCGCGGAAGGATTGGCCACCTTCTTTCGAGAGCAGAGTTGGGTGGCTTGGCTAAG 1139
 |||||
 Db 1084 GAGTGACTAATGACTCACGAAATTTCTCATGAGGATGAATGAGCGATATGGCAAGT 1143
 |||||
 QY 1140 AGCTATCCAGCGGATTTGGCATCTTACTATCCGATGTGACCTTGCAGACCGCAAAATCAC 1199
 |||||
 Db 1144 AATACAAAGCGCTGCTTAGGGAATCTTCCGCAATATAATACGAAGCCCTAAATAT 1203
 |||||
 QY 1200 CTGAGGCGTGATCTCTTCTTACAGGGGNAACCGTACATGAGTGTGTACCTC 1259
 |||||
 Db 1204 ATCCACTGCTACATTTCTTCTCTCTCCGAAATAACAGCCATATTTAAGTGTGTATCTC 1263
 |||||
 QY 1260 CATACCTTCGAAGC 1273
 |||||
 Db 1264 CACTCAITTTGAAC 1277
 |||||

RESULT 6

ABK15523
 ID ABK15523 standard; DNA; 1598 BP.
 XX
 AC ABK15523;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE DNA encoding dimethylallyltryptophan synthase (DmaW) version #2.
 XX
 KW Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
 KW endophyte; forage; pasture; turf; land reclamation; soil conservation;
 KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
 KW Balansia; Balansiopeis; Echinodothis; Atkinsonella; Myriogenospora;
 KW Neotyphodium; Parepichloe; knockout; antisense technology; gene; ds.
 XX
 OS Neotyphodium coenophialum.
 XX
 PH Key Location/Qualifiers
 FT CDS 25..1511
 FT /*tag= a
 FT /product= "dmaW"
 FT /note= "Dimethylallyltryptophan synthase. This sequence
 contains introns, the positions of which are not detailed
 in the specification"

US6335188-B1.

01-JAN-2002.

03-MAR-2000; 2000US-00518657.

22-MAR-1999; 99US-0125490P.

(KENT) UNIV KENTUCKY RES FOUND.

Schardl CL, Wang J;

WPI; 2002-163205/21.

P-PSDB; AAU76413.

XX New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 PT molecule) from fungi that are symbionts of commercially important
 PT grasses, useful to engineer ergot alkaloid-deficient symbionts.

XX Disclosure; Col 19-20; 16pp; English.

XX The invention describes an isolated dimethylallyltryptophan synthase
 CC (dmaW) nucleic acid (1). (1) is a dmaW sequence from epiphyte fungi
 CC Neotyphodium coenophialum symbiont of commercially significant grasses

CC and is useful for expressing DnaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a DnaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epichloe that lack DnaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of DnaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansia, Balansiospiza,
 CC Echinodothia, Atkinsonella, Myriogonospora, Neotyphodium, and
 CC Parepilichia, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This sequence encodes a
 CC dimethylallyltryptophan synthase (dmat) described in the method of the
 CC invention
 XX
 SQ Sequence 1598 BP; 453 A; 392 C; 330 G; 423 T; 0 U; 0 Other;
 Query Match 12.7%; Score 176.6; DB 6; Length 1598;
 Best Local Similarity 50.7%; Pred. No. 3.7e-45;
 Matches 583; Conservative 0; Mismatches 529; Indels 39; Gaps 5;
 33 GCAGCAACTGTGTCGCAAGCCCTTCTAGTGTCTGAGTCAGTCAGCCCTGAACTCTCGAAC 92
 34 GCAAGACACTCCACCAGGAAGTTTATCAACCCCTCAGCGAACAATTTGACTTTGCCAAC 93
 93 AAGGACACACAAATAGTGTATAGCAGAGCTCCGATGTTGCCACCATGATGGCGGG 152
 94 AATGACAGAGGTATGTTGGCAGCAGCGGCCCAATGTTCCAAAGATATCCAACT 153
 153 GCGCGTATGATGTTCAAGCAGATCAAGTTCTCTGTATGATCCACGTCAGGTTCATATC 212
 154 GCTAATATAGCATTTATGTCAGTATCAACATCTGAGCATTTATAAAGCCATATCAT 213
 213 CCGCGTGTGTCATACCCAGAAAGGTCAGCCCATGACCTGGAAGATCATCTACA 272
 214 CTTTCTGTTGTCATCTCAAGAGTGG---CGAGCGGTGGCTAAGCATCTTACG 270
 273 CGCTTCGGAGTCTCTTTCGAGCTGAGCTTCAATTAATCTCAAAATCACTACTACGGTTTGA 332
 271 AGATACGGAACCCGTTTGAGCTAAGTCTTAATGCTCTGACTCCATAGTTCGGTATACA 330
 333 TTGAGCCCTCGGTTCCCTGACGGGAAAGAGATGATCCATTAACACCCAGGCAATC 392
 331 TAGAGGCTATTAAACCGCAACTGGCAGCCATCTCGATCCGTTCAACACTTTCCGCTATC 390
 393 AGGCTGTTCTCAGGACCTCAAGGCCATGGTTCAGGGCTTGACCTGGAATGGTTCCGAT 452
 391 TGGAGGCTCTAAAAGCTTATAGATTCCAGCCAGGCATAGACCTTCAATGGTTTCC 450
 453 CATTTCACTAAAGATTGGTGGTTTCGGAGGAAGAGGCTCGGACTCTGCTAGATCGAGAT 512
 451 TACTTTAAACAGAGCTTACACTTGACGCAAGAGTCCACGTCACCTGCACTCTCAAAAC 510
 513 ATTGAGATCCCGTCTCTCAGACACAGCAAACTGCGAGCCGATCTGGAGCCATCTGGC 572
 511 TTGCTCAAGAAAGATCAAAATCAAAACAGCTAGCTAGCTTAA---AGTGAC 567
 573 GATATTGCTTGAAGACTTACATCTACCCCGGATCAAGTCGATCGGACCGGACCCCA 632
 568 AAGTTCTGACTCAAGACCTACATCTACCCGGAATTTGAAGTCCGTCGCAACTGTAATCG 627
 633 AAGAGAGACTCATGTTTGAGCAAT---CAAGGCTGCGACAAAGTTTGGCAAGTTGCC 689
 628 GTCCAGGAGCTGTGTTTGGCTCCGTCGCAAGCTAGCGCAGAGCATAGATATCCGT 687
 690 ACTCCACTGGCAATCTCTGAGGAGTTTAT--AGCTGAGCAGACCCACCTCTCTCGG-- 745
 688 CCTGCCTTTGAAATGCTAGAGACTATGTCCAGTCTCGCAATAAAGTCCCTACACGGAT 747

Qy 746 -----CCACTTCTCTCATGCGATTTGGTTCAGCCGCTCC 779
 Db 748 GACAGTCACAATACCTCCATTATCTTACGCGCTTCTCTCTTCCGACCTGGTGAAGTCTTACC 807
 Qy 780 GAGTCCCGAATCAAGGTCTACTGTATGAAAGCCAGCTCGACCTGGCTCCATCGAAGGT 839
 Db 808 AAGTCTCGTCAAGATCTACCTCTGGAACGAATGTCTCGTTGCCAGCATGGAAGAT 867
 Qy 840 ATTTGGAATCTCAACGCGGCGACGGAACGATCCAGAGACACTGGAATGTCTGGATGCGCTG 899
 Db 868 CTTTGGAGCGCTTGGCGCGCTTCGAAAGATCAGTCCACTATTGAGGATTTGGAGATGATC 927
 Qy 900 AGGAGCTGTGGAGCTATTGCCCGTCAAGAGGCTGTGTCTCACTGCGCACTGCTTT 959
 Db 928 CGAGAATCTTGGGTCTCTTACATGTCCTCTGTTGGCGCTTACCTCGACCTTAC 987
 Qy 960 TACGAGCGGCTACTCTACCGCAGGAGCAGCTCCCTTCAATTAATAATTTTACCTTCTCT 1019
 Db 988 TTGCGCTTGGCGCTATTCCTCAATGAGAACTTCCGTCCTGCGCAATTACACCTTACAC 1047
 Qy 1020 CCTAAAAGCGCACTTCCGGAACACAGATCTATTTCCTCTGTTTGGGAGAACGACAAA 1079
 Db 1048 CATAATGATCCGATACCAAGACCGAAAGTGTACTTTACTGTGTTCCGATGATGATG 1107
 Qy 1080 ACCATCGCGAAGGATTGGCCACCTCTTTGAGAGCAGAGGTTGGGCTGGCTTGGCTAAG 1139
 Db 1108 GAGTGACTAATGCACTCAAGAAATTTCTCATGAGGCATGAATGAGCATATGCGCAAGT 1167
 Qy 1140 AGCTATCCAGC 1150
 Db 1168 AATACAAAGC 1178
 RESULT 7
 ID AAF11858 standard; cDNA; 676 BP.
 AC AAF11858;
 XX DT 13-MAR-2001 (first entry)
 DE Aspergillus oryzae EST SEQ ID NO:4381.
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 OS Aspergillus oryzae.
 PN WO200056762-A2.
 XX PD 28-SEP-2000.
 XX PF 22-MAR-2000; 2000WO-US007781.
 XX PR 22-MAR-1999; 99US-00273623.
 XX PA (NOVO) NOVO NORDISK BIOTECH INC.
 XX PI (NOVO) NOVO NORDISK AS.
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 WPI; 2000-594572/56.
 XX Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags.
 XX Claim 88; Page 1875; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring the
 CC global expression of genes from FF cells allows the production potential
 CC of the microorganisms to be improved. New genes may be discovered,
 CC possible functions of unknown open reading frames can be identified and
 CC gene copy number variation and stability can be monitored. The expression
 CC of genes can be used to study how FF cells adapt to changes in culture
 CC conditions, environmental stress, spore morphogenesis, recombination,
 CC metabolic or catabolic pathway engineering. Using ESTs provides several
 CC advantages over genomic or random cDNA clones including elimination of
 CC redundancy as one spot on an array equals one gene or open reading frame,
 CC and organization of the microarrays based on function of the gene
 CC products to facilitate analysis of the results. AAF07478 to AAF11247
 CC represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents
 CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from
 CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from
 CC *Trichoderma reesei*, which are all specifically claimed in the present
 CC invention
 XX
 SQ Sequence 676 BP; 178 A; 156 C; 167 G; 174 T; 0 U; 1 Other;

Query Match 12.6%; Score 175.6; DB 3; Length 676;
 Best Local Similarity 57.8%; Pred. No. 4.7e-45;
 Matches 351; Conservative 0; Mismatches 250; Indels 6; Gaps 2;

QY 52 AGCCCTTCTAGTGGTCTGAGTCCAGCCCTGAACTCTCGAACAGGACACACAAATGCT 111
 DB 70 AACATTCAGGTTCTCAGCAAGATCTTAGATTTTGCACACGAGGACGAGACTGCT 129
 QY 112 GGTATGACAGCTCCGANTTTGACCAACATGATGCGGGGCGGCTGATGTTTCAG 171
 DB 130 GGCACAGTACCGGCCAATGATGTCACAAATCTTAAAGGATGCGGCTACGCAATCATG 189
 QY 172 CACAGTACAAAGTTCTCTGTATCCACCGTCAAGTCAATCCCGGGTGGTCCATACC 231
 DB 190 CCATATCTATATCTCTGTCTCCACACAAATGTTGTTCCGTATCTGGGCCCTATC 249
 QY 232 CAGAAAGGCTCAGCCCATGCACTGGAAGATCATCTCACAGCTTCGGACTTCCTTTCG 291
 DB 250 CGGAAACGCGACAGAC---CGATGATGAGCATCTCAGCCGATTCGGTCTACCTTATG 306
 QY 292 AGCTGAGCTCAATTAATCTCAATCACTACTAGGTTTCGATCGAGCCCTCGGTTCGC 351
 DB 307 AATTGAGCTTGAATGCTTAATTCGGTCTGGGTTTCATTCGAGCCCATTCGACCTT 366
 QY 352 TGACGGGACCAAGAGTATGATTCATTCAACACCGAGCAATCAGGCTGTCTCCAGGACC 411
 DB 367 TATCAGGACTGAGCAGACCCCTTCAATGCACATGTCATCTGGGATCTCTTGGGAAC 426
 QY 412 TCAAGCCCATGTTTCAGGGCTGACCTGGAATGTTTCATCAATTCATTAAGCAATGG 471
 DB 427 TCGGAGCTGGGTTCTGATTTTGTATCTCCAGTGGTTTCCCAATTCAGAGGATCTGG 486
 QY 472 TCGTTTCGAGGAGAGGCTCGGACTCTGCTAGATCGAGATATGAGATCCCGCTTCA 531
 DB 487 TTTAGTATCAGAGGAAACGAGTTTGTAGAGACACACGCTCGACAGGACAGGTTA 546
 QY 532 AGACAGACAAACTGGCAGCGGATCTCGAGCCATCTCGGCAATATGTTCTTGAAGACT 591
 DB 547 AGACTCAGAACAACTANGCGTGGACTTGAA---AGGTGGAATAATTCGAGGTCAAGATG 603
 QY 592 ACATCTACCGGGATCAAGTGCATCGGACCGGACCCCAAGAGAGACTCATGTTTG 651
 DB 604 ACATGATCTTATCTGAAATGGTTGCTACCGGATTCCAATTGAGGACTCATGTTTG 663
 QY 652 AGCAAT 658

DB 664 ATTCGAT 670

RESULT 8
 ABK15522
 ID ABK15522 standard; DNA; 1908 BP.
 XX
 AC ABK15522;
 DT 08-MAY-2002 (first entry)
 XX
 DE DNA encoding dimethylallyltryptophan synthase (DmaW) version #1.
 XX
 KW Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
 KW endophyte; forage; pasture; turf; land reclamation; soil conservation;
 KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
 KW Balansia; Balansioopsis; Echinodthis; Atkinsonella; Myriogenospora;
 KW Neotyphodium; Parepichloe; knock out; antisense technology; gene; ds.
 XX
 OS Neotyphodium coenophialum.
 XX
 FH Key Location/Qualifiers
 CDS 277..1756
 FT /*tag= a
 FT /product= "dmaW"
 FT /note= "Dimethylallyltryptophan synthase. This sequence
 contains introns, the position of which are not detailed
 in the specification"
 XX
 PN US6335188-B1.
 PD 01-JAN-2002.
 XX
 PF 03-MAR-2000; 2000US-00518657.
 XX
 PR 22-MAR-1999; 99US-0125490P.
 XX
 PA (KENT) UNIV KENTUCKY RES FOUND.
 XX
 PI Schardl CL, Wang J;
 XX
 DR WPI; 2002-163205/21.
 DR P-ESDB; AAU76412.
 XX
 XX New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 molecule) from fungi that are symbionts of commercially important
 grasses, useful to engineer ergot alkaloid-deficient symbionts.
 XX
 PS Disclosure; Col 17-20; 16pp; English.
 XX
 CC The invention describes an isolated dimethylallyltryptophan synthase
 CC (DmaW) nucleic acid (I). (I) is a DmaW sequence from epiphyte fungi
 CC Neotyphodium coenophialum symbiont of commercially significant grasses
 CC and is useful for expressing DmaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a DmaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epichloe that lack DmaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of DmaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansia, Balansioopsis,
 CC Echinodthis, Atkinsonella, Myriogenospora, Neotyphodium, and
 CC Parepichloe, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This sequence encodes a
 CC dimethylallyltryptophan synthase (dmaW) described in the method of the
 CC invention

SQ Sequence 1908 BP; 544 A; 449 C; 395 G; 520 T; 0 U; 0 Other;

Query Match 12.4%; Score 173.4; DB 6; Length 1908;
 Best Local Similarity 50.5%; Pred. No. 4.3e-44;
 Matches 581; Conservative 0; Mismatches 531; Indels 39; Gaps 5;

QY 33 GCAGCAACACTGCTGCGCAAGCCCTTCTAGCTGCTGAGTCAAGCCCTGAACTCTCGAAC 92
 DB 286 GCAAGACACTCCACAGAGAGTTATCATACCTTAGCGAACAATTGACTTTGGCAAC 345
 QY 93 AAGGACACACAAAAATGGTGGTATAGACACAGCTCGATTTGGCCACCATGATGGCGGG 152
 DB 346 AATGACAGAGGCTATGGTGGACAGCAGCGGGGCCAATGTCGAAAAGATGCTCCAACT 405
 QY 153 GCGGGTATGATGTTACGACAGATCAAGTTCTCTGTATCCACCGTGAGTCAATC 212
 DB 406 GCTAACTATAGCAATTGATGCTAGTATCGACATCTGGGCAATTTAAGAGCCATGTCAIT 465
 QY 213 CCGCGCTGGGTCCATACCCAGAAAAAGGCTCAGCCCATGCATCGGAGAGTCAATCTCACA 272
 DB 466 CTTTCTTGTGTCTATCTCTACAGAAGTGG--CGAGCGGTGGCTAAGCATCTTACG 522
 QY 273 CGTTCGGACTTCCTTCGAGCTGAGCTTCAATTAATCAATCACTACTAGCTTGA 332
 DB 523 AGATACGGAAACCCGTTTGAAGTAACTTAAATGCTCGGACTCGTAGTTCGGTATACA 582
 QY 333 TTCGAGCCCTCGGTTCCCTGAGCGGAACGAGAGATGATCCATTCACACCCAGGCAATC 392
 DB 583 TAGAGCCTATTAAAGCGCAACTGCGAGTCTATCGATCGTTTAACTTTGCTATC 642
 QY 393 AGCGCTGTTCTCAGGACCTCAAGGCCATGTTTCCAGGCTTGAACCTGGAATGTTGAT 452
 DB 643 TGGGAGGCCCTGAAAAGCATATTGAGTCCAGCCAGGCATAGACCTTGAATGGTTTCT 702
 QY 453 CATTTCACTAAAGCATTTGCTGTTTCGGAGAGAGGCTCGGACTCTGCTAGATCGAGAT 512
 DB 703 TACTTTAAAGAGCTTACACTTGAAGCAAGTCCAGTCACTGACTCGCAAAAC 762
 QY 513 ATTGAGATCCCGTCTTCAAGACACAGAACTCGCAGCCGATCTGGAGCCATCTGCG 572
 DB 763 TTGTTAAGGAACAGATCAAACTCAAACTCAAACTCGTTTGGACCTTAA---AGTGAC 819
 QY 573 GATATGCTTTGAAGACCTACATCTACCCGGATCAAGTCAATCGGACCGGACCCCA 532
 DB 820 AAGTTCGTACTGAAGACCTACATCTACCCGGAATGAAGTCCGTCGCAACTGGTAAATCG 879
 QY 633 AAAGAGAGACTCATGTTTGAAGCAAT---CAAGGCTGCGCAAGTTTGGCAAGTTGCG 689
 DB 880 GTCCAGGAGCTCGTGTGCTCGTCCGCAAGTACGCGCAGAGCAAGATATCCGT 939
 QY 690 ACTCCACTGGCAATCTCGAGGAGTTTAT--AGTGAAGAGCAGCCACCCCTCTCGG-- 745
 DB 940 CCGTCCCTTTGAAATGCTAGAAGACTATGTCCAGTCTCGCAATAAATTTCTTACCAGGAT 999
 QY 746 -----CCACTTTCTCTCATGCAATTTGGTCAAGCCGCTCG 779
 DB 1000 GACAGTCAAACTACTCTATTATCTTACGCTTCTCTTGGACCTGATAGTCTTACC 1059
 QY 780 GAGTCCCGAATCAAGGTTACTGATGGAACGCGAGCTGACCTGCGCTCCATCGAAGT 839
 DB 1060 AAGTTCGTGTCAAGATCTACCTCTGGAACGAATGCTCGTTCGAGCGATGGAAGAT 1119
 QY 840 ATTGAGCTCTAACGCGCAAGAGATCCAGAGACTGATGCTGAGTGGCTG 899
 DB 1120 CTTTGGACGCTTGGCGGCCGTCGAGAGATCAGTCCACTATTGAGGATTTGGAGATC 1179
 QY 900 AGGAGAGCTGGCAGCTATTGCCCCGTCAGCGAGGCTCTGTGTCACCTGCGCAACTGCTTT 959
 DB 1180 CGAGAACTTTGGGGTCTCTCAACATGCTCTCTGTTTGGCGCTTACCCTGAGCCTTAC 1239
 QY 960 TAGAGCGCGGTACTCAGCGAGGAGAGCTGCCCTCTGATTAATAATTTTACTTGTCT 1019
 DB 1240 TTGCCCTCGCGGCCATTCCCAATGAGCAACTTCCGTCCTTCCATGGCCCAATTACACCTTACAC 1299

QY 1020 CCTAAAGGCGACTTCCGACACAGATCTATTTCCTCTGTTTGGGACAGACGACAA 1079
 DB 1300 CATATGATGATCCATACAGAACCGAAGTGTACTTTACTGTGTCGCGCATGATGATG 1359
 QY 1080 ACCATCGCGGAAGATTGGCCACCTTCTTTGAGAGCAGAGGTGGGTGGCTTGGCTAAG 1139
 DB 1360 GAGGTGACTAATGCACTCAGACATTTCTCATGAGGCATGAATGAGCGATATGCGAAGT 1419
 QY 1140 AGCTATCCAGC 1150
 DB 1420 AAATACAAAGC 1430

RESULT 9

AAFL12651
 ID AAFL12651 standard; cDNA; 657 BP.

XX AC AAFL12651;

XX DT 13-MAR-2001 (first entry)

XX DE Aspergillus oryzae EST SEQ ID NO:5174.

XX KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX OS Aspergillus oryzae.

XX PN WO200056762-A2.

XX PD 28-SEP-2000.

XX PF 22-MAR-2000; 2000WO-US007781.

XX PR 22-MAR-1999; 99US-00273623.

XX PA (NOVO) NOVO NORDISK BIOTECH INC.

XX PA (NOVO) NOVO NORDISK AS.

XX PI Berka RM, Rey WM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX DR WPI; 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags.

XX Claim 88; Page 2165; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring the
 CC global expression of genes from FF cells allows the production potential
 CC of the microorganisms to be improved. New genes may be discovered.
 CC possible functions of unknown open reading frames can be identified and
 CC gene copy number variation and stability can be monitored. The expression
 CC of genes can be used to study how FF cells adapt to changes in culture
 CC conditions, environmental stress, spore morphogenesis, recombination,
 CC metabolic or catabolic pathway engineering. Using ESTs provides several
 CC advantages over genomic or random cDNA clones including elimination of
 CC redundancy as one spot on an array equals one gene or open reading frame,
 CC and organisation of the microarrays based on function of the gene
 CC products to facilitate analysis of the results. AAF07478 to AAF11247
 CC represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents

Dd		497	KKRRRWRGMYENRWKYMGARYMTYRCARKKYSISAARCARCWYRGKYYWAGWMW	555
Qy		623	CGGGACCCCAAAAGAGAGACTCATGTTTTGACGAATCAAGGCTGCCGACAAGTTTTGGCAA	682
.	:	:	: :	:
Dd		557	KEYKENWYWKMMWYKKRYSKCSWYCKMSYVASCMKSARKAGAOKCRKSMASWSMRS	616
Qy		693	AGTGGCACCCTCCATCGCAATCTCGAGGAGTTTATAGCTGAGCGAGACCCACCTCCT	742
Dd	:	:	: :	:
Dd		617	SRKRCKASCASSAARYAMGMGTSGSRMRWSYTCYWRKMGSMKSTCTTWYMYMSKYTY	676
Qy	:	:	: :	:
Dd		743	CGGCACATTCTCTCATGCGATTGGTCAAGCCGTC	778
	:	:	: :	:
Dd		677	AKYGSWEYRYRANCMYMWEEVYRYRSYMTYNAWY	712
RESULT 12				
ACA31530		ID	ACA31530 standard; DNA; 1258 BP.	
XX		AC	ACA31530;	
XX		DT	19-JUN-2003 (first entry)	
XX		DE	Prokaryotic essential gene #13187.	
XX		KW	Antisense; ds; prokaryotic essential gene; cell proliferation;	
KM		KM	drug design; gene.	
XX		OS	Enterobacter cloacae.	
XX		PN	WO200277183-A2.	
XX		PD	03-OCT-2002.	
XX		PF	21-MAR-2002; 2002WO-US009107.	
XX		PR	21-MAR-2001; 2001US-00815242.	
PR		PR	06-SEP-2001; 2001US-00948993.	
PR		PR	25-OCT-2001; 2001US-0342923P.	
PR		PR	08-FEB-2002; 2002US-00072851.	
PR		PR	06-MAR-2002; 2002US-0362699P.	
XX		PA	(ELIT-) ELITRA PHARM INC.	
XX		PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;	
PI		PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;	
XX		XX		
WPI:		P-PSDB;	ABU27660.	
DR		DR	WPI: 2003-029926/02.	
DR		XX	New antisense nucleic acids, useful for identifying proteins or screening	
PT		PT	for homologous nucleic acids required for cellular proliferation to	
PT		PT	isolate candidate molecules for rational drug discovery programs.	
XX		XX		
PS		PS	Claim 14; SEQ ID NO 19400; 1766pp; English.	
XX		XX	The invention relates to an isolated nucleic acid comprising any one of	
XX		XX	the 6213 antisense sequences given in the specification where expression	
CC		CC	of the nucleic acid inhibits proliferation of a cell. Also included are:	
CC		CC	(1) a vector comprising a promoter operably linked to the nucleic acid	
CC		CC	encoding a polypeptide whose expression is inhibited by the antisense	
CC		CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated	
CC		CC	polypeptide or its fragment whose expression is inhibited by the	
CC		CC	antisense nucleic acid; (4) an antibody capable of specifically binding	
CC		CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular	
CC		CC	proliferation or the activity of a gene in an operon required for	
CC		CC	proliferation; (7) identifying a compound that influences the activity of	
CC		CC	the gene product or that has an activity against a biological pathway;	
CC		CC	required for proliferation, or that inhibits cellular proliferation; (8)	
CC		CC	identifying a gene required for cellular proliferation or the biological	
CC		CC	pathway in which a proliferation-required gene or its gene product lies	
CC		CC	or a gene on which the test compound that inhibits proliferation of an	
CC		CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a	

compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

XX	
CC	The invention relates to an isolated nucleic acid molecule (I) comprising
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC	specification or its complement. (I) is useful for: (a) assessing whether
CC	a patient is afflicted with prostate cancer; (b) monitoring the
CC	progression of prostate cancer in a patient; (c) assessing the efficacy
CC	of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC	the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;
CC	(f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC	determining whether prostate cancer has metastasized in a patient; (h)
CC	assessing the aggressiveness or indolence of prostate cancer in a patient
CC	; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX	
SO	Sequence 484 BP: 125 A; 106 G; 121 G; 132 T; 0 U; 0 Other;

XX Novel Cynomolgus Monkey Calcium-Sensing Receptor polypeptides are useful
PT for treating infections, pain, cancer, diabetes, obesity, asthma,
PT schizophrenia, hypertension, urinary retention, Parkinson's disease and
PT stroke.
XX
XX
XX Claim 1; Page 27-28; 38pp; English.
XX
CC The sequence represents the coding sequence of Cynomolgus monkey calcium
CC -sensing receptor. The calcium-sensing receptor nucleic acid and
CC polypeptide are useful for treating bacterial, fungal, protozoan and
CC viral infections, particularly infections caused by HIV-1 or HIV-2, pain,
CC cancer, diabetes, obesity, anorexia, bulimia, Parkinson's disease, acute
CC heart failure, hypotension, hypertension, urinary retention,
CC osteoporosis, angina pectoris, myocardial infarction, stroke, ulcers,
CC asthma, allergies, benign prostatic hypertrophy, migraine, vomiting,
CC psychotic and neurological disorders including anxiety, schizophrenia,
CC manic depression, depression, delirium, dementia and severe mental
CC retardation, and dyskinesias, such as Huntington's disease or Gilles de
CC la Tourette's syndrome. The polypeptide is useful for structure-based
CC design of agonist, antagonist or inhibitor of the calcium-sensing
CC receptor. The DNA sequence is useful for chromosome localisation studies.
CC The polypeptide and nucleic acid are also useful as vaccines. (Updated on
CC 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 3237 BP; 746 A; 923 C; 824 G; 744 T; 0 U; 0 Other;

Query Match 2.5%; Score 34.8; DB 4; Length 3237;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 114; Conservative 0; Mismatches 112; Indels 2; Gaps 1;
773 GCGCTCGAGTCCGGAATCAAGGTCTACTGTATGGAACGCGACGTGACCTGCCCTCCAT 832
Db 3009 GCITTTCTGGGCTCCAGGAGTCTGGTGTGTAGATTCCTGTGGGCAATGCGTCTT 2950
833 CGAAGGTATTGAGTCTCAACGGGGACGGAACGATCCAGAGAC--ACTGGATGCTCTG 890
Db 2949 CTGAGGCTCATCGAAGCTCAGTGAGAGGTGACCGTGCCACTGCCGAAGATGACCTTCTG 2890
891 GATGCGCTGAGGAGCTGTGCGAGCTATTGCCCGTCAACGAGGCTGTGTCCACTGCCG 950
Db 2889 CTTCATCTTGGCTGCTGTGCGATTTGCTGTGAGGAGGTGAGGGCTGTGCTG 2830
951 AACTGCTTTTACGACCGCGGTACCTCACGCGAGGAGCTGCCCTTC 998
Db 2829 CTGCTCTTGTGGTGTAGGGCCAGCGCTGCTGTGCTTCTGCTCTC 2782

RESULT 15
ACA42624
ID ACA42624 standard; DNA; 4254 BP.
XX
AC ACA42624;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #24281.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Pseudomonas aeruginosa.
XX
FN W0200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR P-PSDB; ABU38754.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 30494; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4254 BP; 837 A; 1404 C; 1389 G; 624 T; 0 U; 0 Other;

Query Match 2.5%; Score 34.8; DB 7; Length 4254;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 508 GAGATATTGAGATCCCGCTTCTCAAGCACACAGACAACTGGCAGCCGATCTGGAGCCAT 567
Db 1373 GCGCGAGTGGCATCAGCTGTGTCAGAGCCAGACGCCCAACATCGGCTCGGCGCTCA 1432
QY 568 CTGGCGCATATTGCTTTGAAGACCTACATCTACCCCGGATCAAGTCGATCGCGACGGGA 627
Db 1433 ATGGCGTGACTGTCTGTGTCAGAGCCAGACGGTGCCCAATATCGGTTCCGGCGCAAGTGCA 1492
QY 628 CCCCCAAAAGAGAGACTCATGTTTGAAGCAATCAAGGCTGCCGACAAAGTTTGCA 681
Db 1493 TCACCGTTGTGCAAGCCAGAAATGCGCAAAATATCGGTTCCAGGCGCGAGTGCA 1546

Search completed: July 28, 2004, 22:37:15
Job time : 627 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: July 28, 2004, 22:17:55 ; Search time 121 Seconds
(without alignments)
6388.815 Million cell updates/sec
Title: US-10-099-704-1
Perfect score: 1393
Sequence: 1 ccgaagtgtgagcaatggag.....cgaaaaa.....cgaaaaa 1393
Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 682709 seqs, 277475446 residues
Total number of hits satisfying chosen parameters: 1365418
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B COMB.seq:*
5: /cgn2_6/prodata/2/ina/5A COMB.seq:*
6: /cgn2_6/prodata/2/ina/5B COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query % Match Length DB ID Description
1 1393 100.0 1393 4 US-09-472-364-1 Sequence 1, Appli
2 205.2 14.7 1347 4 US-09-518-657-1 Sequence 1, Appli
3 205.2 14.7 1353 4 US-09-518-657-3 Sequence 3, Appli
4 176.6 12.7 1598 4 US-09-518-657-6 Sequence 6, Appli
5 173.4 12.4 1908 4 US-09-518-657-5 Sequence 5, Appli
6 37.8 2.7 2830 2 US-09-010-928B-1 Sequence 1, Appli
7 35.2 2.5 339 4 US-09-821-976-8976 Sequence 8976, Ap
8 34.8 2.5 606 4 US-09-452-991A-10059 Sequence 10059, A
9 34.8 2.5 4248 4 US-09-252-991A-9867 Sequence 9867, Ap
10 33.6 2.4 1560 4 US-07-669-171-1 Sequence 1, Appli
11 32.8 2.4 9098 4 US-09-358-082A-28 Sequence 28, Appli
12 32.8 2.4 4403765 3 US-09-103-840A-2 Sequence 1, Appli
13 32.8 2.4 4411529 3 US-09-103-840A-1 Sequence 1, Appli
14 32.6 2.3 2416 1 US-08-592-126-60 Sequence 60, Appli
15 32.6 2.3 2416 4 US-09-168-595-60 Sequence 60, Appli
16 32.2 2.3 12848 4 US-09-453-702B-252 Sequence 252, App
17 32.2 2.3 26173 4 US-09-453-702B-69 Sequence 69, Appli
18 32.2 2.3 38155 4 US-09-453-702B-79 Sequence 79, Appli
19 32.2 2.3 48908 4 US-09-453-702B-137 Sequence 137, App
20 32 2.3 240 1 US-08-209-747-11 Sequence 11, Appli
21 32 2.3 240 1 US-08-458-298-11 Sequence 11, Appli
22 32 2.3 308 4 US-09-621-976-80 Sequence 80, Appli
23 32 2.3 337 4 US-08-868-452-29 Sequence 29, Appli
24 32 2.3 339 1 US-08-486-057B-1 Sequence 1, Appli
25 32 2.3 339 1 US-08-470-837-29 Sequence 29, Appli
26 32 2.3 339 2 US-08-789-588-1 Sequence 1, Appli
27 32 2.3 339 3 US-09-123-233-1 Sequence 1, Appli

Sequence 3, Appli
Patent No. 5168051
Sequence 220, App
Sequence 28, Appli
Sequence 1376, Ap
Sequence 3, Appli
Sequence 9, Appli
Sequence 735, App
Sequence 909, App
Sequence 10, Appli
Sequence 12, Appli
Sequence 927, App
Sequence 2, Appli
Sequence 1032, Ap
Sequence 2009, Ap
Sequence 1, Appli
Sequence 1065, Ap
Sequence 1264, Ap

ALIGNMENTS

RESULT 1
US-09-472-364-1
; Sequence 1, Application US/09472364
; Patent No. 6383781
; GENERAL INFORMATION:
; APPLICANT: Christensen, Bjorn Eggert
; APPLICANT: Mollgaard, Henrik
; APPLICANT: Kaasgaard, Svend
; APPLICANT: Lehmbeck, Jan
; TITLE OF INVENTION: Methods for producing polypeptides in
; TITLE OF INVENTION: aspergillus mutant cells
; FILE REFERENCE: 4483.200-US
; CURRENT APPLICATION NUMBER: US/09/472.364
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/139,593
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: 60/117,396
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: PA 1999 00745
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: PA 1998 01726
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer 5956
; NAME/KEY: CDS
; LOCATION: (15)...(1328)
US-09-472-364-1

Query Match 100.0%; Score 1393; DB 4; Length 1393;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCACTGTCGCAAGGCCCTTCT 60
Db 1 CGGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCACTGTCGCAAGGCCCTTCT 60
QY 61 AGCTCTGAGTCAAGCCCTGAACTCTCGAAAGGACCACACAAATGGTGTATAGCA 120
Db 61 AGCTCTGAGTCAAGCCCTGAACTCTCGAAAGGACCACACAAATGGTGTATAGCA 120
QY 121 CAGCTCCGATGTTGGCCACCATGATGGCGGGGGCGGCGGTATGATTTCCGCAAGTACA 180
Db 121 CAGCTCCGATGTTGGCCACCATGATGGCGGGGGCGGCGGTATGATTTCCGCAAGTACA 180

QY 181 AGTTCTCTGTATCCACCGTGTAGTCAATCCCGGGCTGGGTGCATATCCAGAAAGG 240
Db 181 AGTTCTCTGTATCCACCGTGTAGTCAATCCCGGGCTGGGTGCATATCCAGAAAGG 240
QY 241 GTACGCCATCACTGGAAGAGTCACTCACAGCTTCGGACTTCCTTTTCGAGCTGAGCT 300
Db 241 GTACGCCATCACTGGAAGAGTCACTCACAGCTTCGGACTTCCTTTTCGAGCTGAGCT 300
QY 301 TCAATTACTCAAAATCACTACTACGGTTTGCATTTCGAGCCCTCGGTTCCCTGACGGAA 360
Db 301 TCAATTACTCAAAATCACTACTACGGTTTGCATTTCGAGCCCTCGGTTCCCTGACGGAA 360
QY 361 CGRAGGATGATCCATTCAACACCCAGGCAATCAGGCTGTTCTCCAGGACCTCAAGCCA 420
Db 361 CGRAGGATGATCCATTCAACACCCAGGCAATCAGGCTGTTCTCCAGGACCTCAAGCCA 420
QY 421 TGGTTCCAGGCTTGACCTGGAATGTTTCGATCATTTTCACTAAGCAATTTGGTTCGG 480
Db 421 TGGTTCCAGGCTTGACCTGGAATGTTTCGATCATTTTCACTAAGCAATTTGGTTCGG 480
QY 481 AGGAGAGGCTCGACTCTGCTAGATCGAGATATTCGATCCCGTTCCTTCAGACACAGA 540
Db 481 AGGAGAGGCTCGACTCTGCTAGATCGAGATATTCGATCCCGTTCCTTCAGACACAGA 540
QY 541 ACAAACCTGGCAGCGGATCTGGAGCCATCTGGCGATATTGCTTGAAGACCTACATCTAC 600
Db 541 ACAAACCTGGCAGCGGATCTGGAGCCATCTGGCGATATTGCTTGAAGACCTACATCTAC 600
QY 601 CGCGGATCAAGTGCATCGCAGCGGACCCCAAGAGAGACTCATGTTTGAAGCAATCA 660
Db 601 CGCGGATCAAGTGCATCGCAGCGGACCCCAAGAGAGACTCATGTTTGAAGCAATCA 660
QY 661 AGGTCCGCAAGATTGGCAAGTTGCAAGTTCCACTCCATCGCAATCTCGAGGAGTTATAG 720
Db 661 AGGTCCGCAAGATTGGCAAGTTGCAAGTTGCAAGTTGCAAGTTGCAAGTTGCAAGTT 720
QY 721 CTGAGGAGACCCACCTCTCGGCCATTTCTCATGCGATTTGGTCAAGCGCTCG 780
Db 721 CTGAGGAGACCCACCTCTCGGCCATTTCTCATGCGATTTGGTCAAGCGCTCG 780
QY 781 AGTCCCGAATCAAGGCTACTGTATGGAAGCGGCTCGACTGCGCTCCATCGAAGGTA 840
Db 781 AGTCCCGAATCAAGGCTACTGTATGGAAGCGGCTCGACTGCGCTCCATCGAAGGTA 840
QY 841 TTTTGGACTCTCAACGGGCGAGGAAAGATCCAGAGACACTGGATGCTGGATGGCTGA 900
Db 841 TTTTGGACTCTCAACGGGCGAGGAAAGATCCAGAGACACTGGATGCTGGATGGCTGA 900
QY 901 GGGAGCTGTGGAGCTATTGCGCGTCAAGAGGCTGTGTCACTGCGGAACTGCTTTT 960
Db 901 GGGAGCTGTGGAGCTATTGCGCGTCAAGAGGCTGTGTCACTGCGGAACTGCTTTT 960
QY 961 ACCAGCGGGTACTCTCAACGGGCGAGGAAAGATCCAGAGACACTGGATGCTGGATGGCTGA 1020
Db 961 ACCAGCGGGTACTCTCAACGGGCGAGGAAAGATCCAGAGACACTGGATGCTGGATGGCTGA 1020
QY 1021 CTAAAGCGGACTTCCCGAACCACAGATCTATTTCCCTGCTTTGGGCAAGAACGCAAAA 1080
Db 1021 CTAAAGCGGACTTCCCGAACCACAGATCTATTTCCCTGCTTTGGGCAAGAACGCAAAA 1080
QY 1081 CCATCCGGAAGGATTTGGCACTCTTTGAGAGCAGAGGTTGGGTTGGCTTAAGA 1140
Db 1081 CCATCCGGAAGGATTTGGCACTCTTTGAGAGCAGAGGTTGGGTTGGCTTAAGA 1140
QY 1141 GCTATCCAGCGGATTTGGCATCTACTATCCGATGTGGACCTGAGACCGCAATCACC 1200
Db 1141 GCTATCCAGCGGATTTGGCATCTACTATCCGATGTGGACCTGAGACCGCAATCACC 1200
QY 1201 TGCAGGCTGGATCTCTTCTCTTCAAGGGGAAACCGTACATGATGTGTACCTCC 1260
Db 1201 TGCAGGCTGGATCTCTTCTCTTCAAGGGGAAACCGTACATGATGTGTACCTCC 1260
QY 1261 ATACCTTGGAAAGCTTCACTGCTGCTGCCAGGAGTGGCTATGTGTCAAGTATGCCACA 1320

Db 1261 ATACCTTGGAAAGCTTCACTGCTGCTGCCAGGAGTGGCTATGTGTCAAGTATGCCACA 1320
QY 1321 ATCCCTTAGGACTAGTTTATCCCTTCATCTATGATCGTTCGTTGAATGTGTTGGTGA 1380
Db 1321 ATCCCTTAGGACTAGTTTATCCCTTCATCTATGATCGTTCGTTGAATGTGTTGGTGA 1380
QY 1381 AAAAAA 1393
Db 1381 AAAAAA 1393
RESULT 2
US-09-518-657-1
; Sequence 1, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; APPLICANT: Wang, Jinghong
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; TITLE OF INVENTION: Which Encode Therefor and Related Methods
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; EARLIER FILING DATE: 2000-03-03
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Neotyphodium coenophialum
US-09-518-657-1
Query Match 14.7%; Score 205.2; DB 4; Length 1347;
Best Local Similarity 50.8%; Pred. No. 6.5e-57;
Matches 647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;
QY 33 GCAGCAACACTGCTGCCAAGCCCTTCTACGTCTGAGTCAAGCCCTGACCTCTCGAAC 92
Db 10 GCAAGACACTCCACCAGGAAGTTTATCATACCTTAGCGAACAATTGACTTTGCCAAC 69
QY 93 AAGACACACAAATGGTGTATAGCACAGTCCGATTTTCCACCATGATGGCGGG 152
Db 70 AATGACAGAGGCTATGGTGCACAGCACGGCCCAATGTTGCAAAAGATGCTCCAACT 129
QY 153 GCGGCTATGATGTTACGCACAGATCAAGTTCTCTGTATCCACCGTGAAGTCAATC 212
Db 130 GCTAATATAGCATTTGATGCTCAGATTCGACATCTGGGCATTTTAAAGGCCATGCT 189
QY 213 CCGGCTTGGGTCCATACCCAGAAAGGGTCAAGCCATGACCTGGAAGAGTCAATCTCACA 272
Db 190 CTTTTCTTGTGTCTATCTACAGAGTGG---CGAGCGGTGGCTAAGCATTTCTACG 246
QY 273 CGCTTCGAGCTTCTTCGAGCTGAGCTTCAATTAATCTCAATCACTACTACGGTTTGA 332
Db 247 AGATACGGAACCCCGTTTGGCTTAAGTCTTAATGCTCGGACTCCGTTAGTTCGGTATA 306
QY 333 TTCGAGCCCTCGGTTCCCTGACGGGAAAGAGGATGATCCATTCAACACCCAGGCAATC 392
Db 307 TAGGAGCTATTAAACCGCGCACTGGCAGTCTCTGGATCCGTTTAACTTTTCGCTATC 366
QY 393 AGGCTGTTCTCAGGACCTCAAGGCCATGGTTCCAGGGCTTGACCTGGAATGGTTGAT 452
Db 367 TGGAGGCCCTGAAAGAGCATATTGAGTCCAGCCAGGCATAGACCTTGAATGGTTTCT 426
QY 453 CATTTCACTAAAGCATTTGGTCTGTTTCGGAGGAGAGGCTCGGACTCTGTAGATCGAGAT 512
Db 427 TACTTTAAACAGAGCTTACATTTGACGMAAGAGTCACTGACCTGACTCGCAAAAC 486
QY 513 ATTGAGATCCCGTCTTCAAGACACAGAAACAACTCGCAGCCGATCTGGAGCCATCTGCG 572
Db 487 TTGGTTAAGGAAACAGATCAAAACTCAAAACAGCTCGCTTTGGACCTTAA---AGGTGAC 543

Qy	573	GATATTGTC	TTGAAGAC	CTACATCT	ACCGGGG	GNATCA	AGTCG	ATCGAT	CGGAC	CGGAC	CCCCA	632
Db	544	AAGTTCGT	ACTAGAC	CTCATCT	ACCCGAA	TTGAAGT	CCGTCG	CAACTG	TGTAAT	CG		603
Qy	633	AAAGAGAG	ACTCATG	TTTGAC	GCAGCA	AT--CA	AGGCTC	CCGAC	AGTTGG	CAAGT	TGC	689
Db	604	GTCCAGAG	CTCGTGT	TGGCTCG	TCCGNA	GTAGT	CGCAGA	GCACA	AGATAT	CCGT		663
Qy	690	ACTCCA	CTGGCA	ATCTCG	AGGATTTAT	--AG	GTAGC	GAGC	ACCA	CCCTCT	CG--	745
Db	664	CTGCGCTT	TGAAATG	CTAGAGA	CTATGT	CCAGT	CTCGCA	TAAATCT	CTAC	CCCGAT		723
Qy	746	-----	-----	-----	-----	CC	ACTTCT	CTCAT	CGGATTT	TGGTCA	AGCGCTC	779
Db	724	GACAGT	CACAA	TACTTAT	TATCTT	CAGCC	TCTCT	CTTGG	CACCTGA	TAA	GTCCTTACC	783
Qy	780	GAGTCC	CGGAAT	CAAGGT	CTACTG	TATGGA	ACCG	CAGCTCG	CACCTTGG	CGCTCC	ATCGAAGT	839
Db	784	AAGTCT	CGTGTCA	AGATCT	ACTCT	CTGGA	ACGA	TGTCTCG	TGTGCC	AGCGATG	GAAGAT	843
Qy	840	ATTGG	ACTCTCA	ACGGG	CAGCA	GCATC	CAGAG	ACATG	BGATGT	CTGGA	TGCGCTG	899
Db	844	CTTTGG	ACGCTT	TGCGG	CCCTCG	GAGA	AGATCAG	TCCACT	ATTGAG	GGAATTGG	AGATGATC	903
Qy	900	AGG	GAGCTGTGG	CAGCTATT	TGCCCGT	CACG	GAGG	GTCTGT	CTCCAC	TGCCG	AACTGCTTT	959
Db	904	CGAGAA	CTTTGG	GGTCTCT	CAACAT	GTCTCT	GTGTTG	CGCGCT	ACCTG	AGCCTTAC		963
Qy	960	TAC	AGCCGG	GTACTCA	CCGAG	GAGAG	AGTCC	CCCTTCA	TATPAA	ATTTTAC	CTTGCT	1019
Db	964	TTGCC	CTCGG	CCATTTC	CCAATG	AGCA	ATCC	CGTCCAT	TGGCA	ATTAC	ACCTTAC	1023
Qy	1020	CTTAA	AGCG	CACPTCC	CGAAC	CACAG	ATCTAT	TTCCTG	TTTGG	CAGAG	ACGACAA	1079
Db	1024	CAT	ATGATCC	ATAC	CAGAA	ACCG	CAAGT	GTACTTT	GTGTC	GCATGA	TGATG	1083
Qy	1080	ACC	ATCCGGA	AGGATTTGG	CCACTT	TTGAG	ACAG	AGTGGGG	TGGGTGG	CTTGG	CTTAAG	1139
Db	1084	GAG	GTACTAATG	CAC	TACAG	CA	TCTT	CATGAG	GCATGA	ATGAGCG	TATATGGCAAGT	1143
Qy	1140	AG	TATCC	AGGGA	TTGG	CA	TCTACT	ATCC	CGATGTG	ACCTG	CAGACGGCA	1199
Db	1144	AA	ATACAA	AGCTG	CGCTCAG	GGAATCTT	CCCG	CATCAT	GATTAC	GAA	GCCTGAA	1203
Qy	1200	CTG	CAGCG	GTGAT	CTCTCT	CTTACA	AGGG	AAAAA	CCGTAC	ATGAGT	GTGTAC	1259
Db	1204	AT	CACTGT	GCATTTCT	TCTCTCT	CTAC	GAA	AAGAAC	CA	CCCATAT	TAA	1263
Qy	1260	CAT	ACCTT	CGA	AGC							
Db	1264	CAC	TCTTT	GAA	C							

```

RESULT 3
US-09-518-657-3
; Sequence 3, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; APPLICANT: Wang, Jinghong
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; TITLE OF INVENTION: Which Encode Therefor and Related Methods
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; CURRENT FILING DATE: 2000-03-03
; EARLIER APPLICATION NUMBER: 60/125,490
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1353
; TYPE: DNA

```

i. ORGANISM: *Neotyphodium coenophialum*
 us-09-518-657-3

	Query Match	14.7%	Score 205.2;	DB 4;	Length 1353;
	Best Local Similarity	50.8%;	Pred. No. 6.5e-57;		
	Matches 647;	Conservative	0;	Mismatches 588;	Indels 39;
					Gaps
Qy	33	GCAGCACACTGCTCCGAAGCCCTTCTAGTCTGAGTCAAGCCCTGAACTCTCGAAC	92		
Db	10	GCAAGACACTCCACAGGAAGTTTATCAACCTCAGCGAAACATTTGACTTTGCGAAC	69		
Qy	93	AAGGACCACACAAAATGGTGGTATAGCACAGCTCCGATGTTGCCACCATGATGGGGGG	152		
Db	70	AATGACCAGAGGCTATGGTGGCACAGCAGCGGCCAATGTTCCAAAGATACTCCAAACT	129		
Qy	153	GCGGGTATGATGTTACGCGACAGTACAAGTTCCTCTGTATCCACCGTAGGTCATCATC	212		
Db	130	GCTAACTATAGCATTTATGCTCAGTATCAACATCTGAGCATTTTATAAAGCCATATCAT	189		
Qy	213	CCGGGTTGGTCCATATACCCAGAAAAGGTCAGCCCATGCACGTGGAGAGTCACTCACA	272		
Db	190	CTTTTCTGGTGTATCTCTACAGAGTGG---CGAGCGTGGCTAAGCATCTTTACG	246		
Qy	273	CGCTTGGACTTCCTTTGAGCTGAGCTTCAATTACTCCAAATCACTACTACGGTTTGA	332		
Db	247	AGATACGGAACCCGTTTGAGCTTAAGTCTTAATTGCTCTGACTCCATAGTTCGGTATAC	306		
Qy	333	TTGAGGCCCTCGGTTCCCTGACGGGAACGAGGATGATCCATTCAACACCCAGGCAATC	392		
Db	307	TAGAGGCTATTAAACGCCGCACTGGGAGGCATCTGGATCCGTTTCAACATTTTCGGTATC	366		
Qy	393	AGGCTGTTCTCCAGGACCTCAAGGCCATGGTTCCAGGGCTTGACCTTGAATGGTTGAT	452		
Db	367	TGGAGGCTCTAAAAAGCTTATAGATTCCAGCCAGGCATAGACCTTCAATGGTTTTC	426		
Qy	453	CATTCTACTAAGATGTGCTGTTTCGGAGGAGAGGCTCGGACTCTGCTAGATCGAGAT	512		
Db	427	TACTTTAAACAAGAGCTTACACTTGACGCAACGAGTCCACGTACCTGGCACTCTCAAAAC	486		
Qy	513	ATTGAGATCCCCTCTTCAAGACACAGAACAAACTGCGAGCCGATCTGGAGCCATCTGGC	572		
Db	487	TTGCTCAAGAAACAGATCAAACTCAAAACAGCTAGCGTTAGACCTTAA---AGTGAC	543		
Qy	573	GATATTGCTTGAAGACCTTACATCTACCGCGGATCAAGTCGATCGACCGGACCCCA	632		
Db	544	AAGTTCGTACTCAAGACCTACATCTACCCCGAATTGAAAGTCCGTCGCAACTGTAATCG	603		
Qy	633	AAAGAGAGACTCATGTTTGAACGCAAT---CAAGGCTCCGACAAAGTTTGCAGAGTGGC	689		
Db	604	GTCAGAGGCTTGTTTTTGGCTCCGTCGCAAGCTAGCGCAGAAGCATAGAGTATCCGT	663		
Qy	690	ACTCCAATGCAATCTCCAGGAGTTTAT--AGCTGAGCGAGACCCACCCCTCTCGG--	745		
Db	664	CCTGCCTTTGAATGCTAGAAGATATGTCCAGTCTCGCAATAAAGTCCCTACACCGAT	723		
Qy	746	-----CCACTTTCTCATGCGATTGGTTCAGGCGGTCC	779		
Db	724	GACAGTCACAATATCTCCATTAATCTTCAAGGCTTCTCTTTGCGACCTGGTAGTCTCTACC	783		
Qy	780	GAGTCCCGAATCAAGGTTACTGCTATGGAACGCCAGCTGCACTGGCTCCATCGAAGGT	839		
Db	784	AAGTCTCGTCAAGATCTACTCTCTGGAACGAAATGGTCTCGTTGGCAGCGGATGGAAGAT	843		
Qy	840	ATTTGGATCTCAACCGGGGACGGAACGATCCAGACACACTGGATGGTCTGGATCGCGCTG	899		
Db	844	CTTTGGACGCTTGGCGGCGCTCGAGAAGATCAGTCCACTATTGAGGGAATGGAGATGATC	903		
Qy	900	AGGAGCTGTGGCAGTATTGCCGCTCAGGAGGCTCTGTGTCCACTGCGCACTCTT	959		
Db	904	CGAAGACTTTGGGGTCTCTTAAACATGTCTCTTGGTTTGGCGGCTTACCTTGACCTTAC	963		
Qy	960	TACGAGCCGGGTACCTCAACGCGAGGAGAGCTCCCTCTCATTTATAAATTTTACCTTGCT	1019		

Db 964 TTGCCCCCTGGCGCCATTCCCAATGAGCAACTTCCGTCCTATGCCCAATTACACCTTTACAC 1023
 Qy 1020 CCTAAAGCCGACCTTCCGAGACACAGATCTATTTCCTGCTTTTGGGCGAGACGACAA 1079
 Db 1024 CATATGATCCGATACAGAACCGAAGTGTACTTTACTGTGTTCCGATGAATGATG 1083
 Qy 1080 ACCATCCGGAAGATGGCCACTTTCTTTAGAGCAGAGGTTGGGTTGGCTTGGCTAAG 1139
 Db 1084 GAGTGACTAATGCACTACGAAATCTTTCATGAGGCATGAATGGAGCGATATGGCACT 1143
 Qy 1140 AGCTATCCAGCGATTTGGCATCTTACTATCCGATGTCGACCTGAGACCGCAATAC 1199
 Db 1144 AATACAAAGCCGCTTGGGAACTTTCCCGCATATAATACGAAAGCCCTAAATAT 1203
 Qy 1200 CTGAGCGGTGGATCTCTCTTCAAGGGGAAACCGTACATGATGTGTACCTC 1259
 Db 1204 ATCCACTCGTACATTTCTTCTCTACCGAAATTAACAGCCATATTTAAGTGTGTATCTC 1263
 Qy 1260 CATACCTTGAAGC 1273
 Db 1264 CACTCATTTGAAC 1277

RESULT 4
 US-09-518-657-6
 ; Sequence 6, Application US/09518657
 ; Patent No. 6335188
 ; GENERAL INFORMATION:
 ; APPLICANT: Schardl, Christopher L.
 ; APPLICANT: Wang, Jinghong
 ; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
 ; TITLE OF INVENTION: Which Encode Therefor and Related Methods
 ; FILE REFERENCES: P-1060
 ; CURRENT APPLICATION NUMBER: US/09/518,657
 ; EARLIER FILING DATE: 2000-03-03
 ; EARLIER APPLICATION NUMBER: 60/125,490
 ; EARLIER FILING DATE: 1999-03-22
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1598
 ; TYPE: DNA
 ; ORGANISM: Neotrypanium coenophialium
 US-09-518-657-6

Query Match 12.7%; Score 176.6; DB 4; Length 1598;
 Best Local Similarity 50.7%; Pred. No. 1.8e-47;
 Matches 583; Conservative 0; Mismatches 529; Indels 39; Gaps 5;
 Qy 33 GCAGCAACACTGTGCGCAAGCCCTTCTAGTCTGAGTCAAGCCCTGAACTCTCGAAC 92
 Db 34 GCAAGACACTCCACAGGAGTTTATCAACCCCTCAGCGAAACATTTTGACCTTGCAAC 93
 Qy 93 AAGACACACAAATGTTGGTGTATAGCACGCTCCGATGTTTGGCCACCATGATGGGGG 152
 Db 94 AATGACAGAGGTATGTTGGTGGCACAGCGGGCCCAATTTCCAAAGATATCTCCAAAT 153
 Qy 153 GCGGGCTATGATGTTACGCGACAGTCAAGTTCTCTGTATCCACCGTGAAGTCAATCATC 212
 Db 154 GCTAATATAGCATTTATGCTAGTATACATCTGAGCATTTATAAAGCCATATCAT 213
 Qy 213 CCGGCGTGGGTCATACACCAAGAGGTTAGCCCATGATGGAAGAGTCAATCTACA 272
 Db 214 CTTTCTTGGTGTCTATCTCAAGAAAGTGG---CGAGCGGTGGCTAAGCATTTCTTAG 270
 Qy 273 CGCTTCGGACTTCTTTCGAGTGGAGCTCAATTTACTCCAAATCACTACTACGTTTGA 332
 Db 271 AGATACGAACCCGTTTGGCTAAGTCTTAATGCTCTGACTCCATGTTCCGATATACA 330
 Qy 333 TTGAGCCCTCGGTTCCCTGACGGAACGAAGGATGATCCATTTCAACACCGGCAATC 392
 Db 331 TAGGAGCCTATTAAACCGGCACTGGCAGCCACTTGGATCCGTTCAACACTTTTCGCTATC 390

Qy 393 AGGCTGTCTTCCAGGACCTCAAGGCCATGTTTCCAGGGCTTGACCTGGAAATGTTTCAT 452
 Db 391 TGGAGGCTCTTAAAAAGCTTATAGATTCCAGCGCCATAGACCTTCAATGGTTTTC 450
 Qy 453 CATTTCAATAAGATTTGGTGGTTTCGAGAGAGAGGCTCGACTCTGCTAGATCGAGAT 512
 Db 451 TACTTTAAACAAGAGCTTACACTTGAACGAAAGAGTCCAGTACCTTGGACTCTTCAAAAC 510
 Qy 513 ATTGAGATCCCGCTTTCAAGACACAGAACTGGCAGCCGATCTTGGAGCCATCTTGGC 572
 Db 511 TTGTCGAAGAACAGATCAAACTCAAAACAAGCTAGCTTAGACCTTAA--AGGTGAC 567
 Qy 573 GATATGTTTGAAGACTTACATCTACCGGGGATCAAGTCAATGCGAGCCGAGACCCCA 632
 Db 568 AAGTTCTGCTCAAGAGCTTACATCTACCCGAAATGAAGTCCGTCGCAACTGGTAAATCG 627
 Qy 633 AAAGAGAGACTCATGTTTGGAGCAAT--CAAGGCTGCCGACAAAGTTTGGCAAGTTGCC 699
 Db 628 GTCCAGGAGCTTGTGTTTGGCTCGTCCGCAAGCTAGGCGAGAGCATAGAGTATCCGT 687
 Qy 690 ACTCCACTGGCAATCTCGAGGAGTTTAT--AGCTGAGCGAGCACCCACCTCTCGG-- 745
 Db 688 CCTGCTTGAATGCTAGAAGACTATGTCAGTCTCGCAATTAAGTCCCTACCAAGAT 747
 Qy 746 -----CCACTTCTCTCATGCGATTTTGGTCAAGCCGTCC 779
 Db 748 GACAGTCAATACTCCATTAATCTTACGGCTTCTCTTGGACCTGGTGGATCTTACC 807
 Qy 780 GAGTCCCGAATCAAGGTCTACTGTATGGAACGCCAGCTCGACCTCGCTCCATCGAAGT 839
 Db 808 AAGTCTCGTCAAGATCTACTCTCTGGAAGATGTTCTCGTTGCCAGCATGGAAGAT 867
 Qy 840 ATTTGGAATCTCAACGGCGAGCGAAGCATCCAGAGACACTGGATGCTCTGGATGCGCTG 899
 Db 868 CTTTGGAGCGCTTGGCGCGCTCGAGAAAGATCAGTCCACTATTGAGGATTTGGAGATGATC 927
 Qy 900 AGGAGCTGTGGCAGCTATTGCGCGCTCACGAGGCTGTGTCCACTGCGCGAACTGCTTT 959
 Db 928 CGAGAACTTTGGGGTCTCTTAACTGTTCTCTGTTTGGCGCTTACCCTAGCCTTAC 987
 Qy 960 TACAGCGGGTACTCTCACGCGAGGAGCAGCTCCCTTCTATTAATAATTTACCTTGTCT 1019
 Db 988 TTGCCCTCGCGGCAATTCGCAATGAGCAACTTCCGTCCTCAATGAGCAATTTACAC 1047
 Qy 1020 CTAAGAGCGCACTTCCCGAAGCACAGATCTATTTCCTCTTTTGGGAGAGACGACAAA 1079
 Db 1048 CATATGATCGATACAGAACCGCAAGTGTACTTTACTGTGTTGGCATGATGATG 1107
 Qy 1080 ACCATCCGGAAGATTTGGCCACTTCTTTGAGAGCAGAGTTGGGTGCTTGGCTTAAG 1139
 Db 1108 GAGTGACTAATGCACTCAGCAAAATTTCTTATGAGGCATGAATGGAGCGATATGGCAAGT 1167
 Qy 1140 AGCTATCCAGC 1150
 Db 1168 AATACAAAGC 1178

RESULT 5
 US-09-518-657-5
 ; Sequence 5, Application US/09518657
 ; Patent No. 6335188
 ; GENERAL INFORMATION:
 ; APPLICANT: Schardl, Christopher L.
 ; APPLICANT: Wang, Jinghong
 ; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
 ; TITLE OF INVENTION: Which Encode Therefor and Related Methods
 ; FILE REFERENCES: P-1060
 ; CURRENT APPLICATION NUMBER: US/09/518,657
 ; CURRENT FILING DATE: 2000-03-03
 ; EARLIER APPLICATION NUMBER: 60/125,490
 ; EARLIER FILING DATE: 1999-03-22
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5
LENGTH: 1908
TYPE: DNA
ORGANISM: Neotyphodium coenophialum
US-09-518-657-5

Query Match 12.4%; Score 173.4; DB 4; Length 1908;
Best Local Similarity 50.5%; Pred. No. 2.3e-46;
Matches 581; Conservative 0; Mismatches 531; Indels 39; Gaps 5;

QY 33 GCAGCAACATGTCGCAAGCCCTTCTAGTCTGAGTCAGCCCTGAACTCTCGAAC 92
DB 286 GCAAGACATCCACCAGGAAGTTTATCATACCTTTAGGGAACATTTGACCTTTGCAAC 345
QY 93 AAGGACACACAAATGGTGTATAGCACAGCTCCGATGTTGCGACCATGATGGGGG 152
DB 346 AATGACAGAGGCTATGGTGGACAGACCGGCCAAATGTTGAAAAGATGCTCCAACT 405
QY 153 GCGGGTATGATGTTCAACGCACAGTAAAGTTCTCTGTATCCACCGTGAGGTGATCATC 212
DB 406 GCTAACTATAGCATGATGCTAGTATCGACATCTGGCATTTTATAAGAGCCATGTCATT 465
QY 213 CCGCGGTTGGTCCATACCCAGAAAGGTCAGCCCATGCACTGGAAGATCATCTACA 272
DB 466 CTTTCTTGGTGTCTATCTCTCAAGAAGTGG---CGAGCGGTGGCTAAGCATTTCTACG 522
QY 273 CGCTTCGGACTTCTTTTCAGCTGAGCTTCAATTAATCTCCAAATCACTACTAGGTTTGA 332
DB 523 AGATACGGAACCCGGTTTGGAGCTAAGTCTTAATGCTCGGACTCCGCTAGTTCGGTATAC 582
QY 333 TTGAGGCCCTCGGTTCCCTGAGCGGAACGAGGATGATCCATTCMAACCCAGGCAATC 392
DB 583 TACGAGCCTATTAAACCCGCACTGCGAGTCTGATCCGTTTAAACATTTTCGCTATC 642
QY 393 AGCGCTGTTCTCCAGGACCTCAAGGCCATGGTTCCAGGCTTCACTCGAATGTTTCGAT 452
DB 643 TGGAGGCCCTGAAAAGCATATTGATCCAGCCAGGATAGACCTTGAATGGTTTCT 702
QY 453 CATTCACATAAGCATTTGTCGTTTCGAGGAAGAGCTCGAGCTCTGTAGATCGAGAT 512
DB 703 TACTTTAAACAGAGCTTACACTTGACGCAACGAGTCCAGCTACCTGCACTCGCAAAAC 762
QY 513 ATTGATCCCCGCTTCAAGACACAGAACTGGGAGCGGATCTGGAGCCATCTGGC 572
DB 763 TTGGTTAAGAAACAGATCAAACTCAAAACAGCTCGCTTTGGACCTTAA---AGTGAC 819
QY 573 GATATTGCTTTGAAGACCTACATCTACCGCGGATCAAGTCGATCGGACCGGACCCCA 632
DB 820 AAGTTCGTACTGAAGACCTACATCTACCCGGAATGGAAGTCCGTCGCACTGGTAAATCG 879
QY 633 AAGAGAGACTCATGTTTGAAGCAAT---CAAGGCTGCGGACAAAGTTTGGCAAGTTGCC 689
DB 880 GTCAGGAGCTCGTGTGTTGGCTCCGTCGCAAGCTAGCGCAGAAAGCAAGAGTATCCGT 939
QY 690 ACTCCACTGGCAATCCTCGAGGAGTTAT--AGCTGAGGAGCAGCCACCTCTCTCGG-- 745
DB 940 CTTGCTTTGAAATGCTAAGACTATGTCAGTCTCGCAATAAATCTCTACCAAGGAT 999
QY 746 -----CCACTTCTCTCATGCAATTTGGTCAAGCGGTCC 779
DB 1000 GACAGTCACAATACTCTATTATCTTCAAGCTTCTCTTGGACCTGTAAAGTCTTACC 1059
QY 780 GAGTCCCGAATCAAGCTCTACTGTATGGAACCGCAGCTCGACTGCGCTCCATCGAAGT 839
DB 1060 AAGTCTCGTGTCAAGATCTACTCTCGGAACGAATGGTCTCGTTCAGGATGGAAGAT 1119
QY 840 ATTGGACTCTCAACGGGCGGACGGAACGATCCAGAGACACTGGATGGTCTGGATCGGTG 899
DB 1120 CTTTGGACGCTTGGGCGCGCTCGAGAAGATCAGTCCACTATTGAGGATTTGAGATGATC 1179
QY 900 AGGAGCTGTGGCAGATATGCCCTGACGAGGGTCTGTGTCCTACTGCGCAACTGCTTT 959
DB 1180 CGAGAACTTTGGGGTCTCTCAACATGCTCTCTGGTTTGGCGGCTTACCTGAGCCTTAC 1239

QY 960 TACGAGCCGGGTACTCTACCCGAGGAGGAGCTCCCTTCATTATATAATTTTACCTGTCT 1019
DB 1240 TTGCCCCCTCGCGCCATTTCCCAATGAGCAACTTCCCTCCATGAGCAATTTACCTTAC 1299
QY 1020 CTTAAAGCGCAGCTTCCGAAACCAAGATCTATTTCCTGCTTTTGGCAGAAACGACAAA 1079
DB 1300 CATATGATCCATACAGAACCCGCAAGTGTACTTTACTGTGTTCCGCAATGAATGATG 1359
QY 1080 ACCATCGCGGAAGATTGGCCACCTTTTGAAGCAGAGGTTGGGTGGCTTGGCTAAG 1139
DB 1360 GAGGTGATTAATGCACTCAGCAGATTTCTTCATGAGGATGAATGAGCGGATATGCGAAGT 1419
QY 1140 AGCTATCCAGC 1150
DB 1420 AATACAAAGC 1430

RESULT 6

US-09-010-928B-1
Sequence 1: Application US/09010928B
Patent No. 5994099
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V
APPLICANT: Hayashi, Cheryl Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
TITLE OF INVENTION: CODING THEREFOR
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 GATEHOUSE RD. SUITE 500E
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: -
LOCATION: 1..2830 /note= "Flagelliform DNA sequence
OTHER INFORMATION: taken from the 5' region. The putative start codon is at
OTHER INFORMATION: position 219"
FEATURE:
NAME/KEY: CDS
LOCATION: 219..2830
US-09-010-928B-1

Query Match 2.7%; Score 37.8; DB 2; Length 2830;
Best Local Similarity 52.9%; Pred. No. 0.11; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 435 GACCTGGAATGGTTCGATCATTTTCACTAAAGCATGCTCGTTTCGGAGGAGGCTCGG 494

Db 1906 GACCTGGTAGCTTTGGACAGGAGCGTAGGACCTGGTCCGCTCCGACAGGAGCGGTAG 1965
Qy 495 ACTCTGCTAGATCGAGATATTGAGATCCCGCTCTTCAAGACACAGACAACTGCGACCC 554
Db 1966 GATCTGCTGGCTCCGACAGGAGGAGTAGACCTAGTGGCTCCGACAGGAGTGGCGTAG 2025
Qy 555 GATCTGAGGACATCTCGGATATTGCTTTGAAG 587
Db 2026 GAATGAGGCGTAGACACCGGTGGTGTGGAG 2058

RESULT 7
US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8976

Query Match 2.5%; Score 35.2; DB 4; Length 399;
Best Local Similarity 13.4%; Pred. No. 0.24;
Matches 49; Conservative 147; Mismatches 170; Indels 0; Gaps 0;
Qy 802 GTATGGAAGCCAGCTCGACCTGGCTCCATCGAAGTATTGGACTCTCAACGGCGAC 861
Db 377 GTATGAGCATCAATCAGCAAGKKGKWSGAMKRWTKGKYYWYNTSRSYRRYKTSAM 318
Qy 862 GGAACATCCAGACACTGGATGCTGTGATCGCTGAGGAGCTGTGGCAGCTATTGC 921
Db 317 MGRAKWKRRKTKTKMYWVGKGGKGGSTYVAMESRRGSTGRWSYRRAMWRGSKSWGGSY 258
Qy 922 CGTACGAGGAGGTGTGTCTCCACTGCCAAGTCTTTTACGAGCGGGTACTCACCGC 981
Db 257 YRMAGYRSSRWSYSAMWRKKTCTCWKGRSSWSRSTGYAWMYKKSMTSRKRWYK 198
Qy 982 AGGACAGCTCCCTTCATTATAAATTTTACCTTGTCTCTAAAGCGCACTTCCGAC 1041
Db 197 RRKWRKCTSTRTCYRGSTYKWKYTKKRWTRWYVYKSYMSKKTWMTAY 138
Qy 1042 CACAGATCTATTCCTGCTTTTGGCAGAACGACAAACCATCCGCGGAGGATTGGCA 1101
Db 137 YWTKRWKWTSTKWTCTWCTWCKCTTYWMAGMTMYRYRYWYAKRWKSKRCTWSTTCY 78
Qy 1102 CCTCTTTGAGACAGAGGTGGGTGGCTTGCTAAGAGCTATCCAGCGGATTGGCAT 1161
Db 77 KYWAKCWSVWSNMKWSKWSMWKWTYYYYYMMKWSKMTYWSMSYARKCWRTYAK 18
Qy 1162 CCTACT 1167
Db 17 TYTWMT 12

RESULT 8
US-09-252-991A-10059/c
; Sequence 10059, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10059
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10059
Query Match 2.5%; Score 34.8; DB 4; Length 606;
Best Local Similarity 50.0%; Pred. No. 0.41;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 508 GAGATATTGATCCCGCTCTTCAAGACACAGACAACTGGCAGCGGATCTCGAGCCAT 567
Db 466 GCGGAGTGATCATCAGCGCTGTGCAGAGCCAGAACAGCCCAATATCGGCTCGGGGTCA 407
Qy 568 CTGGCGATATTGCTTGAAGACCTACATCTACCCGGGATCAAGTCGATCGCGACCGGGA 627
Db 406 ATGGCGTACTGTCTGCAGAGCAGACCGTGCCTATATCGGTTCTGGCGCAAGTGGA 347
Qy 628 CCCAAAAGAGAGACTCATGTTTGCAGCAATCAAGGCTCCGACAAAGTTGGCA 681
Db 346 TCACCGTTGTGCAAGCCAGATGGCGCAATATCGGTTTCAGGCGGAGTGGA 293

RESULT 9
US-09-252-991A-9867
; Sequence 9867, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9867
; LENGTH: 4248
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9867

Query Match 2.5%; Score 34.8; DB 4; Length 4248;
Best Local Similarity 50.0%; Pred. No. 1.4;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 508 GAGATATTGATCCCGCTCTTCAAGACACAGACAACTGGCAGCGGATCTCGAGCCAT 567
Db 1367 GCGGAGTGATCATCAGCGCTGTGCAGAGCCAGAACAGCCCAATATCGGCTCGGGGTCA 1426
Qy 568 CTGGCGATATTGCTTGAAGACCTACATCTACCCGGGATCAAGTCGATCGCGACCGGGA 627
Db 1427 ATGGCGTGACTGTCTGTGCAGAGCAGAACCGTGTGCAATATCGGTTCTGGCGCAAGTGGA 1486
Qy 628 CCCAAAAGAGAGACTCATGTTTGCAGCAATCAAGGCTCCGACAAAGTTGGCA 681
Db 1487 TCACCGTTGTGCAAGCCAGATGGCGCAATATCGGTTTCAGGCGGAGTGGA 1540

RESULT 10
US-07-669-171-1
; Sequence 1, Application US/07669171
; Patent No. 5304541

GENERAL INFORMATION:
APPLICANT: PURCHIO, ANTHONY F.
APPLICANT: MADISEN, LINDA
APPLICANT: MEWIN, JUNE RAE
TITLE OF INVENTION: TGP-B1/D2: A NOVEL CHIMERIC TRANSFORMING
TITLE OF INVENTION: GROWTH FACTOR-BETA
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/669,171
FILING DATE: 19910314
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-159-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1560 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 261...1430
US-07-669-171-1

Query Match 2.4%; Score 33.6; DB 1; Length 1560;
Best Local Similarity 48.4%; Pred. No. 1.8;
Matches 93; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 812 CCAGCTCGACCTGGCTCCATCGAAGGTATTGGACTCTCAACGGCGGACGGAACGATCC 871
DB 1217 CAACTTCTGCTGGGCGCTGTCCCTACATTTGGAGCTGGACACGCGAGTACAGCAAGT 1276
QY 872 AGAGACACTGGTGTCTGATCGCTGAGGAGCTGTGCAGCTATTGCCCGTCAACGA 931
DB 1277 CTGGCCCTGTACCAACAGCATAAACCGGCGCTCGGCGGCGCGTGTGCTGCGCGCA 1336
QY 932 GGGTCTGTGTCTCACTGCCGAACCTGCTTTACGAGCGGGGTACCTCACCGCAGGAGCAGT 991
DB 1337 GCGGCTGGAGCACTGCCCATCTGTACTAGTGGCGCGCAAGCCCAAGGTGGAGCAGT 1396
QY 992 CCCCTTCATTAT 1003
DB 1397 GTCCAAACATGAT 1408

RESULT 11
US-09-358-082A-28
Sequence 28, Application US/09358082A
Patent No. 6589784
GENERAL INFORMATION:
APPLICANT: Antoniou, Michael
APPLICANT: Crombie, Robert
TITLE OF INVENTION: A Polynucleotide
FILE REFERENCE: CAC00056

CURRENT APPLICATION NUMBER: US/09/358,082A
CURRENT FILING DATE: 1999-07-21
PRIOR APPLICATION NUMBER: GB 9815879.3
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: US 60/107688
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: GB 9906712.6
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/127410
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: GB 9909494.8
PRIOR FILING DATE: 1999-04-23
PRIOR APPLICATION NUMBER: US 60/134016
PRIOR FILING DATE: 1999-05-12
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 28
LENGTH: 9098
TYPE: DNA
ORGANISM: Homo sapiens
US-09-358-082A-28

Query Match 2.4%; Score 32.8; DB 4; Length 9098;
Best Local Similarity 55.2%; Pred. No. 9.8;
Matches 64; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 1278 AGTCTGCTGCCAGGAGGTGGCTATGTGTACGATGGCCACATCCTTAGGACTAGTTT 1337
DB 13 AGTCAAGCCACAGGAGCTGGGATGAGGATATATACAGGCATTCGTTAATGCTGATT 72
QY 1338 ATCCCTTCATCTATCATCGTTGAATGTGTGTCGAAAAAAGAAAAA 1393
DB 73 GTTCTTATCTCTATCTATCTGAGTGTTTCACAAAAAAGAAAAA 128

RESULT 12
US-09-103-840A-2/c
Sequence 2; Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.4%; Score 32.8; DB 3; Length 4403765;
Best Local Similarity 49.4%; Pred. No. 1.2e+02;
Matches 85; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 699 GCAATCTCGAGGAGTTTATAGCTGAGGAGCACCCTCTCTCGGCGCCTTCTCTCA 758
DB 291460 GCAACCATCGCGGAGGTTCGCTCGGACGCGCGCCACCTGTCGCATCGATGTGGAG 291401
QY 759 TGGGATTGGTCAAGCGCTCCGAGTCCCGAATCAAGTCTACTGTATGGAACGCCAGCTC 818
DB 291400 TCCGCGCGGAGAACCTGGCCGAAACCCAGCAGCAAGGTGGCGCGCACCGCGCTGTGGCTC 291341

/ INFORMATION FOR SEQ ID NO: 60:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 2416 base pairs
 / TYPE: nucleic acid
 / STRANDEDNESS: double
 / TOPOLOGY: linear
 / MOLECULE TYPE: cDNA to mRNA
 / HYPOTHETICAL: NO
 / ANTI-SENSE: NO
 / ORIGINAL SOURCE:
 / INDIVIDUAL ISOLATE: TcB.seq
 / US-09-168-595-60

Query Match	2.3%; Score 32.6; DB 4; Length 2416;
Best Local Similarity	47.0%; Pred. No. 5.1;
Matches 101; Conservative 0; Mismatches 114; Indels 0; Gaps 0;	
QY	39 ACACGTGTCGCAAAAGCCCTTCTACGTGCTGAGTCAAGCCCTGAACCTCTCGAACCAAGAC 98
DB	152 ACTCACCGGCCAAGGGCTTCGACGCCCATGCACCTCTGATGCAGTCAGAGAGTA 211
QY	99 CACACAAATGGTGGTATAGCACAGCTCCGATGTTGGCCACCATGATGGCGGGCGCCGC 158
DB	212 GAGGACAGTGGCGGGCAGCGCGATCTGTGATTTGGGTCTGGCCCTAGCTACTTACCAC 271
QY	159 TATGATGTTTCAGGCACAGTACAAGTTCTCTGTATCCACGTTGAGTCAATCATCCCGCG 218
DB	272 TACTATGATGATGCCCGACCATGTACAGGTGTTCCGCCCTGGGCTTAGCATCTCAGG 331
QY	219 TTGGGTCCATACCAGAAAGGGTCAGCCCATGCA 253
DB	332 AATGGGCCCTGTCTTTGGTTTTCAGGAAGCCCTAAGCA 366

Search completed: July 29, 2004, 01:19:52
Job time : 143 secs

OM nucleic - nucleic search, using sw model

Run on: July 29, 2004, 00:10:36 ; Search time 687 Seconds
(without alignments)
9941.884 Million cell updates/sec

Title: US-10-099-704-1
Perfect score: 1393
Sequence: 1 ccgaagtgcaatgag.....cgaaaaaaaaaaaaaa 1393

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:**

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1393	100.0	1393	14	US-10-099-704-1
2	58.2	4.2	992	16	Sequence 1, Appli
3	48.2	3.5	749	16	Sequence 36657, A
4	43.6	3.1	1292	16	Sequence 36210, A
5	38.6	2.8	1184	13	Sequence 37019, A
6	38.6	2.8	1184	15	Sequence 394, App
7	38.6	2.8	1184	15	Sequence 394, App
8	38.6	2.8	1184	15	Sequence 394, App
9	38.6	2.8	1184	15	Sequence 394, App
10	38.6	2.8	1184	15	Sequence 394, App
11	38.6	2.8	1184	15	Sequence 394, App
12	38.6	2.8	1184	16	Sequence 394, App
13	38.6	2.8	1184	16	Sequence 394, App
14	38.6	2.8	1184	16	Sequence 394, App

C 15	38.6	2.8	1184	16	US-10-141-759-394	Sequence 394, App
C 16	38.6	2.8	1184	16	US-10-140-803-394	Sequence 394, App
C 17	38.6	2.8	1184	16	US-10-140-864-394	Sequence 394, App
C 18	37.2	2.7	2289	17	US-10-437-963-71390	Sequence 71390, A
C 19	37.2	2.7	2686	17	US-10-437-963-71401	Sequence 71401, A
C 20	36.4	2.6	66499	17	US-10-322-281-39	Sequence 39, Appl
C 21	35.8	2.6	13288	13	US-10-332-288-27	Sequence 27, Appl
C 22	35.6	2.6	1258	13	US-10-282-122A-19400	Sequence 19400, A
C 23	35.4	2.5	945	16	US-10-260-238-4003	Sequence 4003, Ap
C 24	34.8	2.5	3237	9	US-09-727-205-1	Sequence 1, Appli
C 25	34.8	2.5	4254	13	US-10-282-122A-30494	Sequence 30494, A
C 26	34.6	2.5	1995	16	US-10-369-493-37566	Sequence 37566, A
C 27	34.4	2.5	1065	13	US-10-451-207-10	Sequence 10, Appl
C 28	34.4	2.5	2454	15	US-10-156-761-3622	Sequence 3622, Ap
C 29	34.4	2.5	9025608	15	US-10-156-761-1	Sequence 1, Appli
C 30	34.2	2.5	571	13	US-10-142-426-542	Sequence 542, App
C 31	34.2	2.5	571	15	US-10-123-155-542	Sequence 542, App
C 32	34.2	2.5	571	15	US-10-146-731-542	Sequence 542, App
C 33	34.2	2.5	571	15	US-10-140-473-542	Sequence 542, App
C 34	34.2	2.5	571	15	US-10-141-761-542	Sequence 542, App
C 35	34.2	2.5	571	15	US-10-142-885-542	Sequence 542, App
C 36	34.2	2.5	571	15	US-10-158-790-542	Sequence 542, App
C 37	34.2	2.5	571	16	US-10-137-871-542	Sequence 542, App
C 38	34.2	2.5	571	16	US-10-140-923-542	Sequence 542, App
C 39	34.2	2.5	571	16	US-10-141-759-542	Sequence 542, App
C 40	34.2	2.5	571	16	US-10-141-759-542	Sequence 542, App
C 41	34.2	2.5	571	16	US-10-140-805-542	Sequence 542, App
C 42	34.2	2.5	571	16	US-10-140-864-542	Sequence 542, App
C 43	34.2	2.5	624	13	US-10-027-632-189298	Sequence 189298,
C 44	34.2	2.5	624	16	US-10-027-632-189298	Sequence 189298,
C 45	34.2	2.5	4674	17	US-10-437-963-69259	Sequence 69259, A

ALIGNMENTS

RESULT 1
US-10-099-704-1
; Sequence 1, Application US/10099704
; Publication No. US20020197682A1
; GENERAL INFORMATION:
; APPLICANT: Christensen, Bjorn Eggert
; APPLICANT: Mollgaard, Henrix
; APPLICANT: Kaasgaard, Svend
; APPLICANT: Lehmebeck, Jan
; TITLE OF INVENTION: Methods for producing polypeptides in
; FILE REFERENCE: 4483.200-US
; CURRENT APPLICATION NUMBER: US/10/099,704
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/472,364
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/139,593
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: 60/117,396
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: PA 1999 00745
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: PA 1998 01726
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer 5956
; NAME/KEY: CDS
; LOCATION: (15)...(1328)
US-10-099-704-1

Query Match 100.0%; Score 1393; DB 14; Length 1393;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAAGCTGAGCAATGAGATCTTCAAGAAACGACACACACTGCTGCCAAGACCTTCT 60
DB 1 CCGAAGCTGAGCAATGAGATCTTCAAGAAACGACACACACTGCTGCCAAGACCTTCT 60

QY 61 ACFTGCTGAGTCAAGCCCTGAACCTCTCGAACAAGGACACACAAATGCTGTATAGCA 120
DB 61 ACFTGCTGAGTCAAGCCCTGAACCTCTCGAACAAGGACACACAAATGCTGTATAGCA 120

QY 121 CAGTCCGATGTTGCCACCATGATGCGGGGCGGCTATGATGTTCCAGCACATACA 180
DB 121 CAGTCCGATGTTGCCACCATGATGCGGGGCGGCTATGATGTTCCAGCACATACA 180

QY 181 AGTTCCTCTGTATCCACCGTGAAGTCAATCCCGCGTGGGTCATACCCCAAGAAAGG 240
DB 181 AGTTCCTCTGTATCCACCGTGAAGTCAATCCCGCGTGGGTCATACCCCAAGAAAGG 240

QY 241 GTGAGCCCATGCACTGGAAGTCACTCAACGCTTCGACATTCCTTCGAGCTGAGCT 300
DB 241 GTGAGCCCATGCACTGGAAGTCACTCAACGCTTCGACATTCCTTCGAGCTGAGCT 300

QY 301 TCAATTACTCCAAATCACTACTAGGTTGCTATCGAGCCCTCGGTTCCCTGACGGGAA 360
DB 301 TCAATTACTCCAAATCACTACTAGGTTGCTATCGAGCCCTCGGTTCCCTGACGGGAA 360

QY 361 CGAAGGATGATCAATCAACACCCAGCAATCAGGCTGTCTTCCAGGACCTCAAGGCCA 420
DB 361 CGAAGGATGATCAATCAACACCCAGCAATCAGGCTGTCTTCCAGGACCTCAAGGCCA 420

QY 421 TGGTTCAGGCTTGACCTGGAATGCTCGATCATTTCAATCAAGCATTTGGTTCGGTGG 480
DB 421 TGGTTCAGGCTTGACCTGGAATGCTCGATCATTTCAATCAAGCATTTGGTTCGGTGG 480

QY 481 AGAAGAGGCTCGGACTCTGCTAGATCGAGATATGAGATCCCGCTTTCACAGACACAGA 540
DB 481 AGAAGAGGCTCGGACTCTGCTAGATCGAGATATGAGATCCCGCTTTCACAGACACAGA 540

QY 541 ACAACTGGCAGCGATCTGGAGCCATCTGGGATATGCTTGAAGACCTACATCTACC 600
DB 541 ACAACTGGCAGCGATCTGGAGCCATCTGGGATATGCTTGAAGACCTACATCTACC 600

QY 601 CGCGGATCAAGTCAAGTCCGACCCGAGACCCCAAGAGAGACTCATGTTTGACGCAATCA 660
DB 601 CGCGGATCAAGTCAAGTCCGACCCGAGACCCCAAGAGAGACTCATGTTTGACGCAATCA 660

QY 661 AGCTCCGGAACAAGTTGGCAAGTTGGCCATCTCCACTGCGAATTCCTTCGAGGATTTAG 720
DB 661 AGCTCCGGAACAAGTTGGCAAGTTGGCCATCTCCACTGCGAATTCCTTCGAGGATTTAG 720

QY 721 CTGAGCAGCAGCACCCTCTCGGCACTTCTCTCATGCGATTTGGTCAAGCGCTCG 780
DB 721 CTGAGCAGCAGCACCCTCTCGGCACTTCTCTCATGCGATTTGGTCAAGCGCTCG 780

QY 781 AGTCCCGAATCAAGGTCTACTGTATGGAACGCGAGCTCGACCTCGCTCCATCGAAGGTA 840
DB 781 AGTCCCGAATCAAGGTCTACTGTATGGAACGCGAGCTCGACCTCGCTCCATCGAAGGTA 840

QY 841 TTTGGACTCTCAAGCGGCGAGACGATCCAGAGACCTGATGCTGGATGGCTGCA 900
DB 841 TTTGGACTCTCAAGCGGCGAGACGATCCAGAGACCTGATGCTGGATGGCTGCA 900

QY 901 GGGAGCTGTGGAGCTATTGCCCGTCAAGAGGGTCTGTGTCCATCTGCGGAACTCTTTT 960
DB 901 GGGAGCTGTGGAGCTATTGCCCGTCAAGAGGGTCTGTGTCCATCTGCGGAACTCTTTT 960

QY 961 ACAGCGGGGTACCTCACCGAGGAGCAGCTCCCTCTCATTAATTAATTTACCTTGTCTC 1020
DB 961 ACAGCGGGGTACCTCACCGAGGAGCAGCTCCCTCTCATTAATTAATTTACCTTGTCTC 1020

QY 1021 CTAAAGCGCACTTCCCGAACCACAGATCTATTTCCTGCTTTTGGGCAAGAACGACAAA 1080

DB 1021 CTAAAGCGCACTTCCCGAACCACAGATCTATTTCCTGCTTTTGGGCAAGAACGACAAA 1080

QY 1081 CCATCGCGGAGGATTTGCCACCTTCTTTGAGACAGAGGTTGGGTGGCTTGGCTAAGA 1140
DB 1081 CCATCGCGGAGGATTTGCCACCTTCTTTGAGACAGAGGTTGGGTGGCTTGGCTAAGA 1140

QY 1141 GCTATCCAGCGGATTTGGCATCTTACTATCCCGATGTGGACCTGCAGACCGCAATCACC 1200
DB 1141 GCTATCCAGCGGATTTGGCATCTTACTATCCCGATGTGGACCTGCAGACCGCAATCACC 1200

QY 1201 TGCAGGCGTGTATCTCTTCTTACAAGGGGAAAAACCGTACATAGTGTGTACCTCC 1260
DB 1201 TGCAGGCGTGTATCTCTTCTTACAAGGGGAAAAACCGTACATAGTGTGTACCTCC 1260

QY 1261 ATACCTTCGAAGCGTTCACTGCTGCTGCCAGGAGGTGGCTATGTGTCAAGATGGCCACA 1320
DB 1261 ATACCTTCGAAGCGTTCACTGCTGCTGCCAGGAGGTGGCTATGTGTCAAGATGGCCACA 1320

QY 1321 ATCCTTAGGACTAGTTTATCCCTTCACTTATGCAATCCGTGATGTTGGTTCGAAAAA 1380
DB 1321 ATCCTTAGGACTAGTTTATCCCTTCACTTATGCAATCCGTGATGTTGGTTCGAAAAA 1380

QY 1381 AAAAAA 1393
DB 1381 AAAAAA 1393

RESULT 2
US-10-369-493-36657
; Sequence 36657, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 36657
; LENGTH: 992
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-36657

Query Match 4.2%; Score 58.2; DB 16; Length 992;
Best Local Similarity 48.2%; Pred. No. 1.5e-08;
Matches 259; Conservative 0; Mismatches 268; Indels 10; Gaps 3;

QY 429 GGGCTTACCTGGAATGTTTCGATCATTTTCACTAAAGCATTTGGTTCGTTTCGAGGAGAG 488
DB 403 GGATTCGATATGACATCTTTCGATCATGTTTCGGGATACATCTTTCATTTTCGACGAGGAG 462

QY 489 GCTCGGACTCTGTAGATCGAGATATTGAGATCCCGTCTTCAAGACACAGAACAACTG 548
DB 463 ACAGCTGCTCCCTTGAAGAACCCAGAT-----GCCTATGGGATCAAGACGATATCTGCTCTA 516

QY 549 GCAGCCGATCTGAGCGCATCTGCGGATATTGTCTTGAAGACCTTACATCTACCGCGGATC 608
DB 517 GGATTCGATATGAG---GGGACACACAGGTGGTCAAGTGTCTATATGATCTCGCGGTGG 573

QY 609 AGTTCGATCGGACCGGGGACCCCAAGAGAGACTCATGTTTGAACGCAATCAAGGCTGCC 668
DB 574 AAGAGTCTTTCACACGGGGGTTTCGGTTCGCAAGCTGATACGAGATTCGCTTTCGAGCGGATC 633

QY 669 GACAGTTCGCAAGAGTTGCTCACTCCACTGGCAATCTCTGAGGAGTTTATAGCTGAGCGA 728


```

US-10-142-426-394/c
; Sequence 394, Application US/10142426
; Publication No. US20040048339A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRAN
; FILE REFERENCE: ACIDS ENCODING
; FILE REFERENCE: P330R1C224
; CURRENT APPLICATION NUMBER: US/10/14
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See File
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-426-394

Query Match 2.8%; Score
Best Local Similarity 11.2%; Pred.
Matches 33; Conservative 82; Mi

QY 92 CAAGGACCCACAAAATGGTGGTAT
Db 990 C.MT.RMVCYN.M....C.D.S.

QY 152 GCGCGGTATGATGTTCCACGACAG
Db 930 .Y.CS.M..SBMCN.DY.YY.ATSS
QY 212 CCGCGCGTTGGGTCCATACCAGAA
Db 870 .....N.N.YSTNY.Y.DN...NB
QY 272 ACGCTTCGGACTCTCTTCGAGCTG
Db 810 .HWCT.KKT.MBHMDS.CB.G.C.C
QY 332 ATTGAGGCCCTCGGTTCCTCGAGC
Db 750 RTYNHMV.AYG.SK.T...C.C.

RESULT 6
US-10-123-155-394/c
; Sequence 394, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J

```

```

; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-146-731-394

Query Match
Best Local Similarity 2.8%; Score 38.6; DB 15; Length 1184;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY 92 CAAGGACCACAAAATGGTGTATAGCAGCTCGATGTTGGCCACCATGATCGCGG 151
Db 990 C.MT.RMNCYDN.M.....C.D.S...S...B.SHA...B.NY.T.TTBS.YB.SY.. 931
QY 152 GCGCGCTATGATGTTTCACGACAGTACAAAGTTCCTCTGTATCCACCGTGAAGTCA 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTBM..YSY..CAT 871
QY 212 CCGGCGTGGTCCATACCCAGAAAGGTCAGCCCATGCTGGAAGATCATCTCAC 271
Db 870 .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACGTTCCGAGCTTCCTTTCCAGCTGAGCTTCAATTAATCAATCACTACTACGGTTGC 331
Db 810 .HWCT.KKT.MBHMDS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY 332 ATTGAGCCCTCGGTTCCCTGAGCGGAACGAGGATGATCCATCAACACCCAG 386
Db 750 RTYNHM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

```

```

RESULT 8
US-10-140-472-394/c
; Sequence 394, Application US/10140472
; Publication No. US2003013888A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-472-394

Query Match
Best Local Similarity 2.8%; Score 38.6; DB 15; Length 1184;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

```

```

QY 92 CAAGGACCACAAAATGGTGTATAGCAGCTCGATGTTGGCCACCATGATCGCGG 151

```

```

Db 990 C.MT.RMNCYDN.M.....C.D.S...S...B.SHA...B.NY.T.TTBS.YB.SY.. 931
QY 152 GCGCGCTATGATGTTTCACGACAGTACAAAGTTCCTCTGTATCCACCGTGAAGTCA 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTBM..YSY..CAT 871
QY 212 CCGGCGTGGTCCATACCCAGAAAGGTCAGCCCATGCTGGAAGATCATCTCAC 271
Db 870 .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACGTTCCGAGCTTCCTTTCCAGCTGAGCTTCAATTAATCAATCACTACTACGGTTGC 331
Db 810 .HWCT.KKT.MBHMDS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY 332 ATTGAGCCCTCGGTTCCCTGAGCGGAACGAGGATGATCCATCAACACCCAG 386
Db 750 RTYNHM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

RESULT 9
US-10-141-761-394/c
; Sequence 394, Application US/10141761
; Publication No. US2003014832A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-761-394

Query Match
Best Local Similarity 2.8%; Score 38.6; DB 15; Length 1184;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

```

```

QY 92 CAAGGACCACAAAATGGTGTATAGCAGCTCGATGTTGGCCACCATGATCGCGG 151
Db 990 C.MT.RMNCYDN.M.....C.D.S...S...B.SHA...B.NY.T.TTBS.YB.SY.. 931
QY 152 GCGCGCTATGATGTTTCACGACAGTACAAAGTTCCTCTGTATCCACCGTGAAGTCA 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTBM..YSY..CAT 871
QY 212 CCGGCGTGGTCCATACCCAGAAAGGTCAGCCCATGCTGGAAGATCATCTCAC 271
Db 870 .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACGTTCCGAGCTTCCTTTCCAGCTGAGCTTCAATTAATCAATCACTACTACGGTTGC 331

```

```
Db      810   .HWCT.KKT.MHBHMS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY      332   ATTGAGCCCTCGGTTCCCTGACGGAAGGATGATCCATTCAACACCCAG 386
          :|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      750   RTYNHM.AVG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

RESULT 10
US-10-142-885-394/c
; Sequence 394, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-885-394

Query Match      2.8%; Score 38.6; DB 15; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY      92   CAAGGACCACAAATGGTGTATGACAGCTCGGATGTTGGCACCATGATCGCGG 151
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      990   C.MT.RMNCYDN.M....C.D..S...S...B..SHBA...B.NY.T.TTBS.YB.SY.. 931
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      152   GCGCGCTATGATGTTCCAGCCACAGTACAAAGTTCTCTGATCCCAATCACTACTACGTTGC 211
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      930   Y.CS.M..SBMCN.DY.YY.ATSS.T..Y.T..SC.HCT.YCA.SCTEM..YSY..CAT 871
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      212   CCGCGGTGGTCCATACCCGAAAGGTCAGCCCATGCACTGGAAGAGTCACTCTCAC 271
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      870   .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      272   ACCTTCGGACTTCCTTCGAGCTGAGCTTCAATTACTTCAATCACTACTACGTTGC 331
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      810   .HWCT.KKT.MHBHMS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      332   ATTGAGCCCTCGGTTCCCTGACGGAAGGATGATCCATTCAACACCCAG 386
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      750   RTYNHM.AVG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

RESULT 11
US-10-158-790-394/c
; Sequence 394, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
```

```
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C448
; CURRENT APPLICATION NUMBER: US/10/158,790
; CURRENT FILING DATE: 2002-05-30
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-158-790-394

Query Match      2.8%; Score 38.6; DB 15; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY      92   CAAGGACCACAAATGGTGTATGACAGCTCGGATGTTGGCACCATGATCGCGG 151
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      990   C.MT.RMNCYDN.M....C.D..S...S...B..SHBA...B.NY.T.TTBS.YB.SY.. 931
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      152   GCGCGCTATGATGTTCCAGCCACAGTACAAAGTTCTCTGATCCCAATCACTACTAC 211
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      930   Y.CS.M..SBMCN.DY.YY.ATSS.T..Y.T..SC.HCT.YCA.SCTEM..YSY..CAT 871
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      212   CCGCGGTGGTCCATACCCGAAAGGTCAGCCCATGCACTGGAAGAGTCACTCTCAC 271
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      870   .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      272   ACCTTCGGACTTCCTTCGAGCTGAGCTTCAATTACTTCAATCACTACTACGTTGC 331
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      810   .HWCT.KKT.MHBHMS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      332   ATTGAGCCCTCGGTTCCCTGACGGAAGGATGATCCATTCAACACCCAG 386
          |: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      750   RTYNHM.AVG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

RESULT 12
US-10-137-871-394/c
; Sequence 394, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
```

```

; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-871-394

Query Match          2.8%; Score 38.6; DB 16; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY 92 CAAGGACACACAAAATGGTGGTATAGCAGAGCTCCGATGTTTCCACCACATGATGGGGG 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 990 C.MT.RMNCYN.M....C.D.S...S...B..SHEA...B.NY.T.TTBS.YB.SV.. 931

QY 152 GGGCGGTATGATGTTCAAGCAGCAGTACAAAGTTCCTCTGTATCACCAGTGAATCATCAT 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 930 Y.CS.M..SBMCN.DY.YY.ATSS.T.Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871

QY 212 CCGGCGGTGGTCCATACCAGAAAGGGTCACCCCATGCACATGGAAGAGTCATCTCAC 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 870 .....N.N.YSTNY.Y.DN...NBAH...B.YS.T.TC.NC.NY.AVR.A.A.D..T.T.G 811

QY 272 ACGTTCGGAGCTTCCTTTCGAGCTGAGCTTCAATATCTCCAAATCACTACTACGGTTTGC 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 810 .HWCT.KKT.MHBWDS.CB.G..CT.CY.B.B.D.HTM..DM..YM..VPAS.T.YGN 751

QY 332 ATTCGAGCCCTCGGTTCCCTTCAGCGGAGCAAGAGATGATCCATTCACACCCAG 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 750 RTYNHMM.AVG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSHY.DHBSAB 696

```

```

RESULT 13
US-10-140-923-394/c
Sequence 394, Application US/10140923
Publication No. US20030207355A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C188
CURRENT APPLICATION NUMBER: US/10/140,923
CURRENT FILING DATE: 2002-05-07
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 394
LENGTH: 1184
TYPE: PRT

```

```

; ORGANISM: Homo Sapien
US-10-140-923-394

Query Match      2.8%; Score 38.6; DB 16; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

Qy  92  CAAGGACCACAAAATGFGGTATAGCAGCTCCCATGTTGCCACCATGATGCGGG 151
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  990 C.MT.RMNCYN.M....C.D..S...S...B..SHBA...B.NY.T.TTBS.YB.SY.. 931
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy  152 GCGCGGTATGATGTTCCACGACAGTACAAGTTCTCTGTATCCACCGTGAGGTTCAT 211
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  930 Y.CS.M..SBMCN.DY.YV.ATSS.T.Y..T..SC.HCT.YCA.SCTBM...YSY..CAT 871
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy  212 CCGCGCGTTGGTCCATACCCAGAAAGGTCAGCCCATGCATCGGAAGTATCTCTCAC 271
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  870 .....N.N.YSTNY.Y.DN...NEAH...B.YS.T.TC.NC.NY.AVR.A.A.D..T..G 811
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy  272 ACGTCTGGACTTCCTTCGAGTCGAGCTCAATTACTCCAAATCACTACTACGGTTTGC 331
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  810 .HWCT.KKT..MBHMD.S.CB.G..CT..CY.B.B.D.HTM..DM...Y.M..YRAS.T.YGN 751
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy  332 ATTCGAGCCCTCGGTTCCCTCGAGCGGAAGGATGATGCATTCACACCCAG 386
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  750 RTYNHMM.AVG.SK.T...C.C..BT.G.BA.KSKS..TC.MYTSYV.DHBSAB 696
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-10-141-756-394/c
; Sequence 394, Application US/10141756
; Publication No. US20030207359A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3303RIC200
; CURRENT APPLICATION NUMBER: US/10/141,756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-756-394

```

	Query Match	2.8%;	Score 38.6;	DB 16;	Length 1184;
	Best Local Similarity	11.2%;	Pred.No. 0.052;		
	Matches 33; Conservative	82;	Mismatches 180;	Indels 0;	Gaps 0
Qy	92 CAAGGACACACAATGGTGGTATAGCACAGCTCCGATGTTGCCACCATGATGGCGG	151			
	. : : : :	: : :			
Dd	990 C.MT.RMNCYN.M...C.D.S...S.B..SHEA...B.NY.T.TTBS.YB.SV..	931			
Qy	152 GGCCGGGTATGATGTTTCACGCACAGTACAAAGTTCCTCTGTATCCACCGTGAGTCAATC	211			
	. : : : :	: : :			

Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871
QY 212 CCGGGGTTGGTCCATACCCAGAAAGGTGAGCCCATGCACTGCAAGAGTCACTCAC 271
Db 870N.N.YSTNY.Y.DN..NEAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACGCTTCGAGACTTCCTTCGAGCTGAGCTTCAATTAATCAATCACTACTACGGTTTC 331
Db 810 .HWCT.KKT.MBHMD.S.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY 332 ATTCGAGCCCTCGGTTCCCTGACGGGAGGAGGATGATCCATTCAACACCCAG 386
Db 750 RTYNHMM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSHY.DHBSAB 696

RESULT 15

US-10-141-759-394/c
; Sequence 394, Application US/10141759
; Publication No. US20030207361A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C197
; CURRENT APPLICATION NUMBER: US/10/141,759
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-759-394

Query Match 2.8%; Score 38.6; DB 16; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;
QY 92 CAAGGACACACAAATGGTGTATAGCACAGCTCGGATTTGCCACCATGCGGG 151
Db 990 C.MT.RMNCYDN.M....C.D..S...S...B..SHBA...B.NY.T.TTBS.YB.SY.. 931
QY 152 GCGCGCTATGTTTCAGCGCAGTACAAAGTTCCTCTGTATCCACCGTGGAGTTCATCAT 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871
QY 212 CCGGGGTTGGTCCATACCCAGAAAGGTGAGCCCATGCACTGCAAGAGTCACTCAC 271
Db 870N.N.YSTNY.Y.DN..NEAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACGCTTCGAGACTTCCTTCGAGCTGAGCTTCAATTAATCAATCACTACTACGGTTTC 331
Db 810 .HWCT.KKT.MBHMD.S.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY 332 ATTCGAGCCCTCGGTTCCCTGACGGGAGGAGGATGATCCATTCAACACCCAG 386
Db 750 RTYNHMM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSHY.DHBSAB 696

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 22:11:55 ; Search time 3926 Seconds
(without alignments)
10595.522 Million cell updates/sec

Title: US-10-099-704-1

Perfect score: 1393

Sequence: 1 ccgaagctgagcaatggag.....cgaaaaaaaaaaaaaaa 1393

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_estci:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_est3:*
 - 12: gb_est4:*
 - 13: gb_est5:*
 - 14: gb_est6:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_fam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rod:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	70.8	5.1	557	28	BZ424676
C 2	70.8	5.1	599	28	BZ424729
C 3	70.8	5.1	600	28	BZ424673
C 4	70.8	5.1	602	28	BZ424733

5	46.4	3.3	1201	13	BX381961
6	42.6	3.1	601	28	BZ424674
7	42.6	3.1	601	28	BZ424732
8	41.2	3.0	482	28	BZ424543
9	40	2.9	1019	28	BZ699387
10	39.4	2.8	438	14	CF301206
11	38.4	2.8	602	9	AV609620
12	38.4	2.8	836	10	BE873147
13	38.4	2.8	885	13	BX425603
14	38.4	2.8	1232	12	BM548562
15	38.2	2.7	1190	14	CD508019
16	38.2	2.7	1201	13	BX457923
17	38	2.7	1201	13	BX381961
18	37.8	2.7	397	13	BQ171803
19	37.8	2.7	1201	13	BX382648
20	37.8	2.7	1435	11	BC050967
21	37.4	2.7	361	29	CE628815
22	37.2	2.7	740	13	BU683350
23	37.2	2.7	854	14	CB628664
24	37	2.7	299	9	AV217652
25	37	2.7	906	12	BG298933
26	36.8	2.6	154	12	BU565098
27	36.8	2.6	284	28	BZ424728
28	36.8	2.6	353	28	BZ424731
29	36.8	2.6	479	12	BJ210477
30	36.8	2.6	489	28	BZ424734
31	36.8	2.6	494	28	BZ424678
32	36.8	2.6	541	28	BZ424665
33	36.8	2.6	588	28	BZ424675
34	36.8	2.6	597	13	BU805984
35	36.8	2.6	1201	9	AL543409
36	36.6	2.6	530	28	AQ634647
37	36.6	2.6	645	13	BQ414025
38	36.6	2.6	742	28	AQ752529
39	36.4	2.6	320	9	AV665857
40	36.4	2.6	1016	14	CF266426
41	36.2	2.6	526	9	AL383443
42	36.2	2.6	637	13	BU895985
43	36	2.6	348	12	BM751760
44	36	2.6	588	14	CA814207
45	35.8	2.6	119	12	BU560428

ALIGNMENTS

RESULT 1
BZ424676/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ424676 557 bp DNA linear GSS 13-DEC-2002
100017849-5752 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.

BZ424676.1 GI:26666131

GSS.

Aspergillus terreus

Aspergillus terreus

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

1 (bases 1 to 557)

Askenazi, V., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S.,

Zimmer, D.P., Boers, M.-E., Blomquist, P.R., Martinez, E.J.,

Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K.,

Tobin, J., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C. and

Madden, K.T.

Integrating transcriptional and metabolite profiles to direct the

engineering of lovastatin-producing strains

Unpublished (2002)

Contact: Zimmer DP

Microbia, Inc.

One Kendall Square Building 1400 W, Cambridge, MA 02139, USA

Tel: 617-621-8322

Fax: 617-

Email: dzimmer@microbia.com

Class: plasmid ends
Location/Qualifiers
1. .557
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/notes="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 5.1%; Score 70.8; DB 28; Length 557;
Best Local Similarity 49.1%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

QY 54 CCCTTCTACGCTGCTGAGTCAAGCCCTGAACCTCTCGAACAAGGACACACACAAAATGGTGG 113
DB 453 CCCTGGAGATATTTGGGACAAACGACCGGATTTCCAAACCAAGACGAGGCTCTGGTG 394
QY 114 TATAGCACAGCTCGATGTTTGGCCACCATGATGGCGGGGCGGGCTATGATGTTACGCA 173
DB 393 CTAACACACCGCCCTCTGCTCAACGAATTTCTGGCGAGTGCATATGACGTCACCTTG 334
QY 174 CAGTACAACTCTCTGTATCCACCGTGAAGTCACTCCCGCGTGGTCCATACCA 233
DB 333 CAGTACCAATACCTCAGTTCTTCCGCCACCATGTCATTCCTGTTCTAGGGCCCTCTTT 274
QY 234 GAAAGGGTCAGCCCATGCACTGGAAGAGTCACTCAGACGCTTCGAGCTTCTTTTCGAG 293
DB 273 GCCCAGGAGCACTCCAACTTCGCCAGCAGACTCAGCAAGCAGCCACCTCTTGAT 214
QY 294 CTGAGCTTCAATTT-----ACTCCAAATCACTACTAGGTTTGATTCAGGCCCTCGT 347
DB 213 TTGAGGCTCAATTTCCAGGAGTCCGGTCAACAGTCCGAATGAGCTGGGGGCCATCGT 154
QY 348 TCCCTGACGGGAACGAAGATGATCCATTCACACCCAGGCAATCAGGCTGTCTCCAG 407
DB 153 AGCTTTGCTGCTTCGAGCAGGATCGTTGAACCACTCAGGGCGAGAGTTCTCGAC 94
QY 408 GACCTCAAGCCATGTTTCAGGGCTTGACCTGGAATGTTGATCATTTCACTAAAGCA 467
DB 93 AGCTAGCCATCTTTGACCCACCGTGGACTTGCAGTTGTTCAACACTTCGAGATGAG 34
QY 468 TTGCTGCTTTCCGAGGAAGAGGCTCGGA 495
DB 33 TTGCGCATCAACCATGCCGACGGCTGA 6

RESULT 2
BZ424729/c
LOCUS
100018692-5746 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.

ACCESSION
BZ424729
VERSION
BZ424729.1 GI:26666184
SOURCE
GSS.
ORGANISM
Aspergillus terreus
Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 599)
Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S.,
Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J.,
Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K.,
Tobin, J., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C. and
Madden, K.T.
Integrating transcriptional and metabolite profiles to direct the
engineering of lovastatin-producing strains
Unpublished (2002)

Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
1. .559
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/notes="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 5.1%; Score 70.8; DB 28; Length 599;
Best Local Similarity 49.1%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

QY 54 CCCTTCTACGCTGCTGAGTCAAGCCCTGAACCTCTCGAACAAGGACACACAAAATGGTGG 113
DB 453 CCCTGGAGATATTTGGGACAAACGACCGGATTTCCAAACCAAGACGAGGCTCTGGTG 394
QY 114 TATAGCACAGCTCGATGTTTGGCCACCATGATGGCGGGGCGGGCTATGATGTTACGCA 173
DB 393 CTAACACACCGCCCTCTGCTCAACGAATTTCTGGCGAGTGCATATGACGTCACCTTG 334
QY 174 CAGTACAACTCTCTGTATCCACCGTGAAGTCACTCCCGCGTGGTCCATACCA 233
DB 333 CAGTACCAATACCTCAGTTCTTCCGCCACCATGTCATTCCTGTTCTAGGGCCCTCTTT 274
QY 234 GAAAGGGTCAGCCCATGCACTGGAAGAGTCACTCAGACGCTTCGAGCTTCTTTTCGAG 293
DB 273 GCCCAGGAGCACTCCAACTTCGCCAGCAGACTCAGCAAGCAGCCACCTCTTGAT 214
QY 294 CTGAGCTTCAATTT-----ACTCCAAATCACTACTAGGTTTGATTCAGGCCCTCGT 347
DB 213 TTGAGGCTCAATTTCCAGGAGTCCGGTCAACAGTCCGAATGAGCTGGGGGCCATCGT 154
QY 348 TCCCTGACGGGAACGAAGATGATCCATTCACACCCAGGCAATCAGGCTGTCTCCAG 407
DB 153 AGCTTTGCTGCTTCGAGCAGGATCGTTGAACCACTCAGGGCGAGAGTTCTCGAC 94
QY 408 GACCTCAAGCCATGTTTCAGGGCTTGACCTGGAATGTTGATCATTTCACTAAAGCA 467
DB 93 AGCTAGCCATCTTTGACCCACCGTGGACTTGCAGTTGTTCAACACTTCGAGATGAG 34
QY 468 TTGCTGCTTTCCGAGGAAGAGGCTCGGA 495
DB 33 TTGCGCATCAACCATGCCGACGGCTGA 6

RESULT 3
BZ424673
LOCUS
100017849-5764 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.

ACCESSION
BZ424673
VERSION
BZ424673.1 GI:26666128
SOURCE
GSS.
ORGANISM
Aspergillus terreus
Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 600)
Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S.,
Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J.,
Madden, K.T.

Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and Madden,K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
FEATURES
source
1..600
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 5.1%; Score 70.8; DB 28; Length 600;
Best Local Similarity 49.1%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 54 CCCTTCTACGTCGTGAGTCAAGCCCTGAACCTCTGCAACGAGGACACACAAAATGGTGG 113
DB 71 CCCTGGAAGATATTGGACAAAGACCGGATTCCTCCAAACCAAGACGAGGCTGTGGTGG 130
QY 114 TATAGCAGCTCCGATGTTTGGCCACCATGATGGGGGGGGCGGCTATGATTTCAAGCA 173
DB 131 CTAAACACCGCCCTCTGCTCAACGAATTTCTGGCCGAGTGCAATATGACGTCCACTTG 190
QY 174 CAGTACAGTTCTCTGTATCCACCGTGAGGTATCATCCCGCGTGGTCCATACCCA 233
DB 191 CAGTACCAATCTCAGTTCTTCCGCCACCATGTCATTCCTGTTCTAGGCGCCCTCTTT 250
QY 234 GAAAGGGTTCAGCCCATGCACTGGAAGAGTCTATCTCACAGCTTCGGACTTCCTTTTCGAG 293
DB 251 GCCCCAGGACGACTCCAACTTCGCCAGCAGACTCAGCAAGCAGCGCCACCTCTTGAT 310
QY 294 CTGAGCTTCAATT-----ACTCCAAATCACTACTACGTTTGGATTCGAGCCCTCGGT 347
DB 311 TTGAGGCTCAATTTCAGGAGTCCGGTGCAACAGTCCGAATGAGCCTGGGGCCCATCGGT 370
QY 348 TCCTCAGCGGAACGAGGATGATCCATTCAACACCCAGGCAATCAGGCTCTGTTCTCCAG 407
DB 371 AGCTTGTCTGGCTTCAGCAGAGATCCGTTGAAACAGTTCAGGGCGAGAGATTTCTGAC 430
QY 408 GACCTCAAGGCCATGTTCCAGGGCTTGACCTGGAATGTTTGATCATTTTCACTAAAGCA 467
DB 431 AAGCTAGCCATCTGTATACCAACCGTGGACTTGCAGTTGTTCAACACTTCGAGAATGAG 490
QY 468 TTGGTCTTTCGGAGGAAGGCTCGGA 495
DB 491 TTTGGCATCAACCATGCCAGCGCTGA 518

RESULT 4
BZ424733 602 bp DNA linear GSS 13-DEC-2002
LOCUS BZ424733
DEFINITION 100018692-5758 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.
ACCESSION BZ424733
VERSION BZ424733.1 GI:26666188
KEYWORDS GSS.
SOURCE Aspergillus terreus

ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 602)
AUTHORS Akenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S., Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J., Sykes,K., Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and Madden,K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
FEATURES
source
1..602
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 5.1%; Score 70.8; DB 28; Length 602;
Best Local Similarity 49.1%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 54 CCCTTCTACGTCGTGAGTCAAGCCCTGAACCTCTGCAACGAGGACACACAAAATGGTGG 113
DB 72 CCCTGGAAGATATTGGACAAAGACCGGATTCCTCCAAACCAAGACGAGGCTGTGGTGG 131
QY 114 TATAGCAGCTCCGATGTTTGGCCACCATGATGGGGGGGGCGGCTATGATTTCAAGCA 173
DB 132 CTAAACACCGCCCTCTGCTCAACGAATTTCTGGCCGAGTGCCAATATGACGTCCACTTG 191
QY 174 CAGTACAGTTCTCTGTATCCACCGTGAGTCTATCCCGGCTGGGTTCATACCCA 233
DB 192 CAGTACCAATCTCAGTTCTTCCGCCACCATGTCATTCCTGTTCTAGGCGCCCTCTTT 251
QY 234 GAAAGGGTTCAGCCCATGCACTGGAAGAGTCAATCTCACAGCTTCGAGCTTCCTTTTCGAG 293
DB 252 GCCCCAGGACGACTCCAAACTTCGCCAGCAGACTCAGCAAGCAGCGCCACCTCTTGAT 311
QY 294 CTGAGCTTCAATT-----ACTCCAAATCACTACTAGGTTTGCATTCGAGCCCTCGGT 347
DB 312 TTGAGGCTCAATTTCAGGAGTCCGGTGCAACAGTCCGAATGAGCCTGGGGCCCATCGGT 371
QY 348 TCCTCAGCGGAACGAGGATGATCCATTCAACACCCAGGCAATCAGGCTCTGTTCTCCAG 407
DB 372 AGCTTGTCTGGCTTCAGCAGAGATCCGTTGAACAGTTCAGGGCGAGAGATTTCTGAC 431
QY 408 GACCTCAAGGCCATGTTCCAGGGCTTGACCTGGAATGTTTGATCATTTTCACTAAAGCA 467
DB 432 AAGCTAGCCATCTTGTATACCAACCGTGGACTTCAGTTGTTCAAAACACTTCGAGAATGAG 491
QY 468 TTGGTCTTTCGGAGGAAGGCTCGGA 495
DB 492 TTTGGCATCAACCATGCCAGCGCTGA 519
RESULT 5
BX381961 1201 bp mRNA linear EST 08-MAY-2003
LOCUS BX381961

```
DEFINITION BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cdna
clone CSODI072YF05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:304533007
EST. EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen, Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI072CC03NP1.
FEATURES
source
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_types="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
```

ORIGIN

```
Query Match 3.3%; Score 46.4; DB 13; Length 1201;
Best Local Similarity 4.1%; Pred. No. 22;
Matches 29; Conservative 246; Mismatches 433; Indels 0; Gaps 0;
QY 88 CGACAGAGGACACAAAATGGTGTATAGACAGCTCGATGTTGGCCACCATGATGG 147
DB 381 CNVACKANNKKKKACNANNNNNNNKKMMNNNNKKKKCMNKKMMNNKKMMNNKKKK 440
QY 148 CGGGGCGCGGTATGATGTTTCAGCAGACAGTCAAGTTCCTCTGTATCCACCGTGAGGTCA 207
DB 441 KKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 500
QY 208 TCATCCGGGGTGGGTGCATATCCAGAAAAGGGTCAGCCCATGCACTGGAAGATCATC 267
DB 501 KNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 560
QY 268 TCACAGCTTCGGACTTCCTTTCAGCTCAGCTTCAATTACTCCAAATCACTACTACGCT 327
DB 561 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 620
QY 328 TTGCATTCGAGCCCTCGGTTCCTCGACGGGAACGAAGGATGATCATTCACACCCAGG 387
DB 621 KKKKKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 680
QY 388 CAATCAGGCTGTCTCCAGGACCTCAAGCCATGTTCCAGGGCTTGACCTGGATGGT 447
DB 681 KKKKKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 740
QY 448 TCGATCATTTCAATAAGCATTTGGTGTTCGGAGGAAGAGGCTCGGACTCTGCTAGATC 507
DB 741 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 800
QY 508 GAGATATTGAGATCCCGGTCTTCAAGACACAGAAACAAATCGCAGCCGATCTGGAGCCAT 567
DB 801 CNVCKKKKKKKKKKKCMNKKMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 860
QY 568 CTGGCGATATTGTTCTGAAGACCTACATCTACCCCGGATCAAGTCGATCGCGACCGGGA 627
```

```
DB 861 MNNNNKAMNKEDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 920
QY 628 CCCCAAAAGAGAGACTCATGTTTGAACCAATCAAGGCTGCGGACCAAGTTTGGCAAGTTG 687
DB 921 MNCCKMMCKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 980
QY 688 CCATCCATCGGCAATCTCGAGGAGTTTATAGCTGAGCGAGCAGCACCACCTCTCTCGGCC 747
DB 981 KKVCMYCDXBKCMCKKBMKKVCCKKCMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1040
QY 748 ACTTCTCTCATGCGATTTGGTCAAGCGCTCCGAGTCCCGAATCAAGG 795
DB 1041 MNNCKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1088

RESULT 6
B2424674
LOCUS B2424674
DEFINITION 100017849-5762 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.
ACCESSION B2424674
VERSION B2424674.1 GI:26666129
KEYWORDS GSS.
SOURCE Aspergillus terreus
ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Askenazi M., Boegers M.E., Blomquist P.R., Martinez E.J.,
Zimmer D.P., Feibelman T.P., Mayorga M.E., Maxon M.E., Sykes K.,
Tobin J., Cordero E., Salama S.R., Trueheart J., Royer J.C. and
Madden K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the
engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
FEATURES
source
1. .601
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone
library"
/notes="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI;
Sau3A genomic fragments ligated into BamHI digested
pZEROTM-2"
ORIGIN
Query Match 3.1%; Score 42.6; DB 28; Length 601;
Best Local Similarity 52.5%; Pred. No. 1.5e+02;
Matches 93; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 54 CCCTTCTACCTGCTGAGTCAAGCCCTGACCTCTCGAAGAGGACACACAAATCGTGG 113
DB 384 CCCTGGAAGATATTGGGACAAACGACCGGATTCCCAACCAAGACGAGGAGCTCTCGTGG 443
QY 114 TATAGCACACTCCGATGTTTGCACCATGATCGGGGGCGGCTATGATGTTCAAGCA 173
DB 444 CTAAACACGCCCTCTGCTCAACGATTTCTGCGCGAGTGCAATATGCTCCACTTG 503
QY 174 CAGTACAGTTCTCTGTATCCACCGTGAGGTCAATATCCCGGCGTGGTGGTCCATAC 230
DB 504 CAGTACCAATACCTACGTTCTTCCGCCACCATGATCACTCTCTTAGGGCCCTTC 560
```

AUTHORS Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S., Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J., Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K., Tobin, J., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C. and Madden, K.T.

TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains

JOURNAL COMMENT Unpublished (2002)
Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
1. .482
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

FEATURES source
1. .482
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 3.0%; Score 41.2; DB 28; Length 482;
Best Local Similarity 49.5%; Pred. No. 3.1e+02;
Matches 106; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 282 CTTCCTTCGAGCTGAGCTTCATTAATCTCCAAATCACTACTACCGTTTGCATTCGAGCCC 341
Db 468 CTTGATTTACCGTCATTTCCAGAGTCCGGTCCACAGCTCCGAATGAGCTGGGGCC 409
QY 342 CTCGGTTCCTTCGAGGAAAGAGATGATCATTCAACACCCAGGAATCAGGCTGTT 401
Db 408 ATCGTAGCTTGTGGCTTCAGCAGGATCCGTTGAACCACTTCAGGGCGAGAGATT 349
QY 402 CTCAGGACCTCAAGGCGATGTTCCAGGGCTTCACCTGGAATGTTGATCATTTCACT 461
Db 348 CTCGACAGCTAGCCATCTTGTACCAACCGTGAGCTTGCTGTTCAACACTTCAG 289
QY 462 AAAGCATTTGCTGTTTCGGAGGAAGAGGCTCGGA 495
Db 288 AATGAGTTTGGCATCAACCAATCGCGAGCGCTGA 255

RESULT 9
BZ699387/c
LOCUS BZ699387 1019 bp DNA linear GSS 19-FEB-2003
DEFINITION PUCEG32TD ZM.0.6.1.0 KB Zea mays genomic clone ZMMBTa126F15, genomic survey sequence.
ACCESSION BZ699387
VERSION BZ699387.1 GI:28419234
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1. (bases 1 to 1019)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843

RESULT 7
BZ424732
LOCUS BZ424732 601 bp DNA linear GSS 13-DEC-2002
DEFINITION 100018692-5756 Aspergillus terreus random genomic DNA clone library
ACCESSION BZ424732
VERSION BZ424732.1 GI:26666187
KEYWORDS GSS.
SOURCE Aspergillus terreus
ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus. 1 (bases 1 to 601)
REFERENCE Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S., Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J., Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K., Tobin, J., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C. and Madden, K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
1. .601
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 3.1%; Score 42.6; DB 28; Length 601;
Best Local Similarity 52.5%; Pred. NO. 1.5e+02;
Matches 93; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 54 CCCTTCTACGTGTCGAGTCCCTGACCTTCGACACAGGACCCACACAAATGGTGG 113
Db 383 CCTGGAAGATATTGGACAAACGACCGGATTCGCGAGTGCCAAATGACGTCCACTTG 442
QY 114 TATAGCACAGCTCCGATGTTTGCACCATGATGCGGGGGCGGCTATGATGTTCAACGCA 173
Db 443 CTAACACACGCCCTCTGCTCAAGAAATTCGCGAGTGCCAAATGACGTCCACTTG 502
QY 174 CAGTACAGTTCCTCTGTATCCACCGTGAGGTGATCATCCCGCGCTGGTGCATAC 230
Db 503 CAGTACCAATACCTCAGCTTCTTCGCGCACCATGTGATCTCTGTTAGGGCCCTTC 559

RESULT 8
BZ424543/c
LOCUS BZ424543 482 bp DNA linear GSS 13-DEC-2002
DEFINITION 100014823-2817 Aspergillus terreus random genomic DNA clone library
ACCESSION BZ424543
VERSION BZ424543.1 GI:26665998
KEYWORDS GSS.
SOURCE Aspergillus terreus
ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus. 1 (bases 1 to 482)

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

Location/Qualifiers

FEATURES

source

1. 1019
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM87A126F15"
/clone_lib="ZM 0.6-1.0 KB"
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN

Query Match 2.8%; Score 40; DB 28; Length 1019;
Best Local Similarity 58.3%; Pred. No. 4.7e+02;
Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 1029 GCACCTCCGACACAGATCTATTCCTGCTTTGGGAGAACGACAAACCATCGCG 1088
Db 179 GCACCCACCGAACACAAATCTATTATCCGATTTTCAAAACACGACACAATTTGTAT 120
Qy 1089 GAAGGATGGCCACCTCTTTGAGACGAGAGTTGGGTGGCTTGAAGAGCTATCCA 1148
Db 119 GAACAAATGCCCATCTTTAGTAGTGATGGTGTGGCTTTTCGGGTGCCAATCA 60

RESULT 10

CF301206/c

LOCUS 438 bp mRNA linear EST 15-AUG-2003
DEFINITION 7LEAF--05-P13.b1 Rice leaf plasmid cDNA library II (7LEAF) Oryza
sativa cDNA clone 7LEAF--05-P13, mRNA sequence.

ACCESSION CF301206

VERSION

KEYWORDS

SOURCE

ORGANISM

Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 438)

AUTHORS Kim J.S., Jun K.M., Cheong P.J., Kim M.J., Lee T.H., Shin Y.C.,

Song S.I., Kim J.K., Kim Y.-K. and Nahm B.H.

TITLE Large-scale Sequencing Analysis of Rice ESTs

JOURNAL Unpublished (2003)

COMMENT Contact: Nahm B.H.

of Bioscience and Genomics Institute, Greengene Biotech Inc.; Division

Yongin, Kyeonggi, Korea

Tel: 82 31 330 6193

Fax: 82 31 321 6355

Email: bhnahm@bio.com, bhnahm@bio.myongji.ac.kr.

FEATURES

source

1. 438
/organism="Oryza sativa"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:4530"
/clone="7LEAF--05-P13"
/clone_lib="leaf"
/tissue_type="leaf"
/dev_stage="7 days after germination"
/lab_host="E. coli DH10B"
/clone_lib="Rice leaf plasmid cDNA library II (7LEAF)"
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; mRNA was capped
with oligoribonucleotides and then used as templates for
RT-PCR."

ORIGIN

Query Match 2.8%; Score 39.4; DB 14; Length 438;
Best Local Similarity 67.9%; Pred. No. 7.3e+02;
Matches 55; Conservative 0; Mismatches 26; Indels 0; Gaps 0

Qy 1313 TGGCCACAATCCTTAGGACTAGTTTATCCCTTATCTATCCATCCGTTGATGTTGG 1372
Db 156 TGGCCCAACCCCTGGGAAGGGTGGTCTTTTTTTTAAATTTGAATTTGGTGG 97
Qy 1373 TCGAAAAAATTT 1393
Db 96 GGGAAAAAATTT 76

RESULT 11

AV609620

LOCUS

DEFINITION AV609620 Bos taurus lung fetus Bos taurus cDNA clone EILU025E08 5',
mRNA sequence.

ACCESSION AV609620

VERSION AV609620.1 GI:9745290

KEYWORDS

SOURCE

ORGANISM

Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 602)

AUTHORS Takasuga A., Hirotsune S., Itoh R., Jitohzono A., Suzuki H., Aso H.

and Sugimoto Y.

TITLE Establishment of a high throughput EST sequencing system using

poly(A) tail-removed cDNA libraries and determination of 36,000

bovine ESTs

Nucleic Acids Res. 29 (22), E108 (2001)

JOURNAL

MEDLINE

COMMENT

Contact: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazusuga@ocn.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

FEATURES

source

1. 602
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="EILU025E08"
/tissue_type="lung"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus lung fetus"
/notes="Vector: pZLI; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"

ORIGIN

Query Match 2.8%; Score 38.4; DB 9; Length 602;
Best Local Similarity 47.6%; Pred. No. 1.1e+03;
Matches 111; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
Qy 10 GAGCATGGAGATCTCCAGAAAGACGACACACTGTGCGAAGCCCTTCTAGTGCTGA 69
Db 70 GACCATNGGTGGGCTCCCGCGCGCCCACTGCTCTGCTAGGCATCCCTAGATCGTCG 129
Qy 70 GTCAAGCCCTGAACTCTTCGAACAAGGACACACAAAAATGGTGGTATAGCACAGCTCCGA 129
Db 130 CCTGCGCCCTGGCCCTGGCGGTAGCCCGCGGACGCGGCGCCCTAGGAGGGTGGC 189
Qy 130 TGTTCGCCACATGATGGGGGGGGCGGCTATGATTTTCCGACAGTACAAAGTTCCTCT 189
Db 190 TGCTGGGCGGCTGTATGGAGGCGGACGTCAATGAGGAGGCGGTGCAGAGGCGCTGCTCT 249
Qy 190 GTATCCACCGTGAAGGTTCATCCCGCGCTTGGGTGCTTACATCCCAAGAAAGGGT 242


```

5', mRNA sequence.
ACCESSION BM548562
VERSION BM548562.1 GI:18783222
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 1232)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12733 row: j column: 16
High quality sequence stop: 365.
FEATURES
Location/Qualifiers
1..1232
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5732079"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(destroyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."
ORIGIN
Query Match 2.8%; Score 38.4; DB 12; Length 1232;
Best Local Similarity 47.2%; Pred. No. 9.7e+02;
Matches 117; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 748 ACCTTCCTCATCGGATTGGTCAAGCGGTCGAGTCCGAGTCCGAGATCAAGGTTCTACTGTATGG 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 ACCTGGCCACATCGAGGTGGAGCTGTGGCGGAGGCGCCTTCCGTGGCCTGGGCGCCG 286
QY 808 AACGCCAGCTCGACCTGGGCTCCATCGAAGGTATTGGACTCTCAACGGGCGACGGAACG 867
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 TGTCTCTGTCTCACTTGGCTTCAACACCTGCTGAGTGTGCTCCACGAGGCGCTGGAGC 346
QY 868 ATCCAGAGACATGGATGCTGGATGCGCTGAGGAGCTGTGGCAGCTATTGCCCGTCA 927
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
347 GGCTGGGCTGATGCGGCGGCTGGACCTGGAGGGAAACGACTGGACGAGCTGCCACCGC 406
QY 928 CGGAGAGTCTGTCTCCACTGCGCACTGCTTTTACGAGCGGGTACCTCACCGAGGAGC 987
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
407 CGACCATCTCCGAATCGGGGCGCCATACCCCAACCCACTCGCCCATCGAGGCGCCACCC 466
QY 988 AGCTCCCC 995
Db |||||
467 AACTACCC 474

RESULT 15
CD508019/c
LOCUS
DEFINITION CDA88-E02.3, mRNA sequence.
CD508019
ACCESSION
CD508019
LOCUS
DEFINITION CDA88-E02.3, mRNA sequence.
CD508019
ACCESSION

CD508019.1 GI:31438588
EST.
Gasterosteus aculeatus (three spined stickleback)
Gasterosteus aculeatus
ORGANISM Gasterosteus aculeatus
REFERENCE
1 (bases 1 to 1190)
Kingsley, D.M., Reichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J., and Myers, R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 88
High quality sequence start: 15
High quality sequence stop: 876.
FEATURES
Location/Qualifiers
1..1190
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Galinas river, CA"
/db_xref="taxon:69293"
/clone="CDA88-E02"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-CDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."
ORIGIN
Query Match 2.7%; Score 38.2; DB 14; Length 1190;
Best Local Similarity 53.7%; Pred. No. 1.1e+03;
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1103 CTTCTTTGAGCAGAGGTTGGGCTGGCTTGGCTAAAGAGCTATCCAGCGATTGGCATC 1162
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1005 CTTGCTTTACAGCAGACATGGGAGGTACCGGGACGCGTTGTCCGAGAGTTGGACGC 946
QY 1163 CTACTATCCCGATGTGACCTGCAGACCGCAATACCTGCAGCGGTGATCTCTTC 1222
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
945 CTTGGAAGCGGAGCGGACCGGGCTGACGGCCCGCTGCTGTCAGGCGCTGGAGCCAGGT 886
QY 1223 TTACAGGGGAAAAAACCGTACATGAG 1249
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
885 CAAAGAGAGGTGGAGCCGTACCTGAG 859

Search completed: July 29, 2004, 01:17:22
Job time : 3933 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:33:31 ; Search time 54 Seconds

(without alignments)
2286.542 Million cell updates/sec

Title: US-10-099-704-2

Perfect score: 2322

Sequence: 1 MEISKAAATLLPKPFVLSQ.....PEAFSAAQAEVAMCHDGHNP 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2322	100.0	437	3 AAY96961	Aay96961 A. oryzae
2	907.5	39.1	448	5 AAU76412	Aau76412 Dimethylyla
3	904.5	39.0	450	5 AAU76413	Aau76413 Dimethylyla
4	103.5	4.5	880	6 AAO27492	Aao27492 Balsam pe
5	102	4.4	521	5 AAM50327	Aam50327 Human nuc
6	102	4.4	606	5 AAO17868	Aao17868 Pyrin dom
7	102	4.4	635	6 ABO07113	Ab007113 Novel hum
8	101	4.3	878	4 AAB11500	Aab11500 C. sativu
9	101	4.3	878	4 AAB86036	Aab86036 Cucumbe
10	99.5	4.3	503	7 ADE54304	Ades4304 Rat Prote
11	99.5	4.3	503	7 ADE54292	Ades4292 Rat Prote
12	99.5	4.3	503	7 ADE54296	Ades4296 Rat Prote
13	99.5	4.3	503	7 ADE60548	Ades60548 Rat Prote
14	99.5	4.3	503	7 ADE62266	Ades62266 Rat Prote
15	99.5	4.3	503	7 ADE54300	Ades4300 Rat Prote
16	99.5	4.3	1210	7 ADE08065	Ades08065 Novel pro
17	99.5	4.3	3234	5 ABP70132	Abp70132 Human NOV
18	98.5	4.2	629	4 ABG04518	Abg04518 Novel hum
19	98.5	4.2	769	6 ABU38122	Abu38122 Protein e
20	97	4.2	1939	5 ABG97437	Abg97437 A. orient
21	96.5	4.2	2627	6 ABR39483	Ab39483 L. cuprin
22	96.5	4.2	3333	6 ABR39482	Ab39482 L. cuprin
23	95.5	4.1	1171	4 AAU32421	Aau32421 Novel hum
24	95	4.1	249	6 ABP78985	Abp78985 N. gonorr
25	95	4.1	604	3 AAY57087	Aay57087 Mouse ret

26	94.5	4.1	317	2 AAW17105	Aaw17105 Amino aci
27	94.5	4.1	317	2 AAW77838	Aaw77838 Multi-fun
28	94.5	4.1	320	2 AAW17115	Aaw17115 Amino aci
29	94.5	4.1	320	2 AAW77848	Aaw77848 Multi-fun
30	94.5	4.1	329	2 AAW17132	Aaw17132 Amino aci
31	94.5	4.1	329	2 AAW77898	Aaw77898 Multi-fun
32	94.5	4.1	430	2 AAY35283	Aay35283 Chlamydia
33	94.5	4.1	430	2 ABU26782	Abu26782 Protein e
34	94.5	4.1	453	5 ABP47775	Abp47775 Protein #
35	94.5	4.1	879	4 ABG25755	Abg25755 Novel hum
36	94.5	4.1	879	4 ABG28383	Abg28383 Novel hum
37	94	4.0	3913	6 ABM67350	Abm67350 Photorhab
38	93	4.0	302	2 AAW17104	Aaw17104 Amino aci
39	93	4.0	302	2 AAW77837	Aaw77837 Multi-fun
40	93	4.0	305	2 AAW17114	Aaw17114 Amino aci
41	93	4.0	305	2 AAW77847	Aaw77847 Multi-fun
42	93	4.0	833	6 ABU00599	Abu00599 S. pneumo
43	93	4.0	2492	5 AAE18302	Aae18302 Venezuela
44	92.5	4.0	585	6 ABU46996	Abu46996 Protein e
45	92.5	4.0	686	4 ABG25945	Abg25945 Novel hum

ALIGNMENTS

RESULT 1
AAV96961
ID AAY96961 standard; protein; 437 AA.
XX
AC AAY96961;
XX
DT 31-OCT-2000 (first entry)
XX
DE A. oryzae dimethylallyl-cycloacetyl-L-tryptophan synthase.
XX
KW Dimethylallyl-cycloacetoacetyl-L-tryptophan synthase; DCAT-S; beta-CPA;
KW toxin production; elimination; heterologous protein production;
KW filamentous fungi.
XX
OS Aspergillus oryzae.
XX
FN WO200039322-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-DK000726.
XX
PR 23-DEC-1998; 98DX-00001726.
PR 27-MAY-1999; 99DX-00000745.
(NOVO) NOVO NORDISK AS.
Christensen BE, Mollgaard H, Kaasgaard S, Lehmebeck J;
WPI; 2000-452411/39.
N-PSDB; AAA51712.
Producing a polypeptide of interest such as a hormone or enzyme,
comprising cultivating a mutant of a parent Aspergillus cell which
produces less of at least one toxin of interest compared to the parent
cell under the same conditions.
Claim 34; Page 61-62; 66pp; English.
The A. oryzae dimethylallyl-cycloacetoacetyl-L-tryptophan synthase (DCAT-S) is involved in the synthesis of beta-CPA, from cyclo-acetoacetyl-L-tryptophan and dimethylallylpyrophosphate, by its homology to a dimethylallyltryptophan synthase (DMAT-S) from Claviceps purpurea. Aspergillus host cells having a modification in the DCAT-S gene, leading to reduced or eliminated toxin production, are useful for expression of heterologous polypeptides of interest. CC heterologous polypeptides of interest. CC or eliminated comprise kojic acid, 3-nitropropionic acid, emodin, etc. The DCAT-S gene can be used to identify and disrupt similar genes in

CC other filamentous fungal host strains such as Trichoderma, Penicillium
 CC and Fusarium
 XX Sequence 437 AA;
 SQ Query Match 100.0%; Score 2322; DB 3; Length 437;
 Best Local Similarity 100.0%; Pred. No. 1.3e-214;
 Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEISKKAATLLPPYVLSQALNLSKNDHTKMYSTAPMTWAGAGYDVHAQYKFLCI 60
 DB 1 MEISKKAATLLPPYVLSQALNLSKNDHTKMYSTAPMTWAGAGYDVHAQYKFLCI 60
 QY 61 HREVIIPALGPYKPGQPMHWKSHLTRFGLPFELSFNYKSLRFAPEPLGSLTGTCDP 120
 DB 61 HREVIIPALGPYKPGQPMHWKSHLTRFGLPFELSFNYKSLRFAPEPLGSLTGTCDP 120
 QY 121 FNTQAIRPVLDLKVAMPVGLDLEWDFHFTKALVVSBEAEARTLLDRDIEIPVFTQNKLA 180
 DB 121 FNTQAIRPVLDLKVAMPVGLDLEWDFHFTKALVVSBEAEARTLLDRDIEIPVFTQNKLA 180
 QY 181 DLSPSGDIVLKYIYPRIKSIATGTPKRLMFDIAKAADKFGKVAATPLALEEFIAERAP 240
 DB 181 DLSPSGDIVLKYIYPRIKSIATGTPKRLMFDIAKAADKFGKVAATPLALEEFIAERAP 240
 QY 241 TLGHFLSCLDVKPSRSRIKVCWERQDLASIEGTWTLNGRNDPDLGLDALRELWQ 300
 DB 241 TLGHFLSCLDVKPSRSRIKVCWERQDLASIEGTWTLNGRNDPDLGLDALRELWQ 300
 QY 301 LLPVTEGLCPNCFEYEPGTSPOQLPFIINFTLSPKALPEQIYEPAPGQNDKTIASG 360
 DB 301 LLPVTEGLCPNCFEYEPGTSPOQLPFIINFTLSPKALPEQIYEPAPGQNDKTIASG 360
 QY 361 LATFFESRGWGLAKSPADLASIYYPVDVDTQANTHLOAWISFSYKGGKPYMSYLHTFEA 420
 DB 361 LATFFESRGWGLAKSPADLASIYYPVDVDTQANTHLOAWISFSYKGGKPYMSYLHTFEA 420
 QY 421 FSAAAQEVAMCHDGNP 437
 DB 421 FSAAAQEVAMCHDGNP 437

RESULT 2
 AAU76412
 ID AAU76412 standard; protein; 448 AA.
 XX AAU76412;
 AC AAU76412;
 DT 08-MAY-2002 (first entry)
 XX Dimethylallyltryptophan synthase (DmaW) version #1.
 DE Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
 XX endophyte; forage; pasture; turf; land reclamation; soil conservation;
 KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
 KW Balansia; Balansioopsis; Echinodopsis; Atkinsonella; Myriogenospora;
 KW Neotyphodium; Parepichloe; knockout; antisense technology.
 XX Neotyphodium coenophialum.
 XX US6335188-B1.
 PN 01-JAN-2002.
 XX 03-MAR-2000; 2000US-00518657.
 XX 22-MAR-1999; 99US-0125490P.
 XX (KENT) UNIV KENTUCKY RES FOUND.
 XX Schardl CL, Wang J;
 XX WPI; 2002-163205/21.

DR N-PSDB; ABK15520.
 XX New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 PT molecule) from fungi that are symbionts of commercially important
 PT grasses, useful to engineer ergot alkaloid-deficient symbionts.
 PS Claim 1; Col 11-14; 16pp; English.
 XX The invention describes an isolated dimethylallyltryptophan synthase
 CC (dmaW) nucleic acid (I). (I) is a DmaW sequence from epiphyte fungi
 CC Neotyphodium coenophialum symbiont of commercially significant grasses
 CC and is useful for expressing DmaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a DmaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epichloe that lack DmaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of DmaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansia, Balansioopsis,
 CC Echinodopsis, Atkinsonella, Myriogenospora, Neotyphodium, and
 CC Parepichloe, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This is the amino acid sequence
 CC of dimethylallyltryptophan synthase (dmaW) described in the method of the
 CC invention
 XX Sequence 448 AA;
 SQ Query Match 39.1%; Score 907.5; DB 5; Length 448;
 Best Local Similarity 41.3%; Pred. No. 2.6e-78;
 Matches 184; Conservative 78; Mismatches 157; Indels 27; Gaps 7;

QY 7 AATLLPKFPYVLSQALNLSKNDHTKMYSTAPMTWAGAGYDVHAQYKFLCIHREVII 66
 DB 4 AKTLHQEVYTLTSETFDANNDQRLMWHSTAPMFKMLQNTANYSIDAQYRHGIYKSHVI 63
 QY 67 PALGPYKPGQPMHWKSHLTRFGLPFELSFNYKSLRFAPEPLGSLTGTCDPFTQAI 126
 DB 64 PFLGVYPTSRGE-RWLSILTRYGTPFELSLNCSDSVRYTYEINATGSHLDPFTFAI 122
 QY 127 RPVLQDLKAMVPGLDLEWDFHFTKALVVSBEAEARTLLDRDIEIPVFTQNKLAADLEPSG 186
 DB 123 WEALKKHIESQPGIDLEWDFSYFKQELTLDANESTYLHSONLVKEQIKTNKALDLK--G 180
 QY 187 D-IVLKYIYPRIKSIATGTPKRLMFDIAI-KAADKFGKVAATPLALEEFIAERAP--- 240
 DB 181 DXFVLKYIYPRIKSVATGKSVQELVFGSVKLAQKHSIRPAFEMLEDIVQSRNFSST 240
 QY 241 -----TLL-GHFTSCDLVKPSRSRIKVCWERQDLASIEGTWTLNGRNDPDLGLDA 294
 DB 241 DSHNTLLSRLLSCLDISPTKSRVKYLLERMVSLPAMEDLTLGGRREDQSTIEGLEM 300
 QY 295 LRELWQLLPVTGELCPNCFEYEPGTSPOQLPFIINFTLSPKALPEQIYEPAPGQND 354
 DB 301 IRELWGLNLSVGLRAYPEPLGAIPLNEQLSPSMANYTLHNDPIPEPOVITYVFGMND 360
 QY 355 KTAEGLATFFESRGWGLAKSPADLASIYYPVDVDTQANTHLOAWISFSYKGGKPYMSVY 414
 DB 361 MEVNTALTTFMRHESWSDMSKYKACLRSEFPHHDEALNYIHSYISFSYKKNPKYLSVY 420
 QY 415 LHTFE-----AFSAAAQEV 428
 DB 421 LHSFETGKVPVFPFEGIAFDACRRDL 446

RESULT 3
 AAU76413
 ID AAU76413 standard; protein; 450 AA.

127 RVLQDLKAMVPGDLLEWFDHFTKALVVSBEERARTLLDRDIEIPVFTQNKLAADLEPSG 186
 123 WEALKLIDSPQGDILQWFSYFKQELTDANESTYLHSONLVKEQINTQNKALDLK-G 180
 187 D-IVLKYIYPRKSIATGTPKRLMFDAL-KAADKFGKATPLAILLEEFIAER--APT- 241
 181 DKFVLKYYIPELKSVAIGKSVQELVFGSVKLAQKHSIRPAFEMEDIVQSENKVPPT 240
 242 -----LLGHFLSCDIAVKPESERIKVYCMERQDLASIEGIWTLNGRRNDPFDGLDA 294
 241 DDSHTPLSRLSLCDLVSPKSRVKIYLLERVMVSLPAMEDLWTLGRRDQSTIEGLEM 300
 295 LRELWOLLPVTEGLCPPLNCFYBPGTSPQOLPFIINFTLSPKALPEPOLIYPAFGOND 354
 301 IRELWGLNMSFGLRAYPEPPLPLGALPNBOLPSMANYTLHNDPIPEPOVYFTVFGND 360
 355 KTAIEGLATFFESRGWGLAKSPADIASYYPDVLQTAHQLQAWISFSYKPKPYMSVY 414
 361 MEVTNALTKEFMRHEDMSWKYKACLRSEFPFHNYEALNVIHSYISFSYRNKPKYLSVY 420
 415 LHTFE 419
 421 LHSFE 425

RESULT 4
 RAO27492
 ID AAO27492 standard; protein; 880 AA.
 AC AAO27492;
 XX
 DT 06-NOV-2003 (first entry)
 DE Balsam pear (Momordica charantia) lipoxygenase isozyme 2 protein.
 XX
 KW Lipoxygenase; hydroperoxidation; polyunsaturated fatty acid; plant;
 KW fatty acid metabolite synthesis; signal molecule; growth regulation;
 KW development regulation; plant development; wound response;
 XX genetic mapping; hyperoxidation catalysis; Balsam pear; enzyme.
 OS Momordica charantia.
 XX
 PN US2003074693-A1.
 XX
 PD 17-APR-2003.
 XX
 PF 29-JAN-2002; 2002US-00059909.
 XX
 PR 10-FEB-1999; 99US-0119597P.
 PR 09-FEB-2000; 2000US-00501422.
 XX
 PA (CAHO/) CAHOON E B.
 PA (KINN/) KINNEY A J.
 PA (KLEI/) KLEIN T X.
 PA (LEEJ/) LEE J.
 PA (PEAR/) PEARLSTEIN R W.
 PA (RAFA/) RAFALSKI J A.
 PA (SHEN/) SHEN J B.
 PA (THOR/) THORPE C J.
 PA (TING/) TINGEV S V.
 PA (WENG/) WENG Z.
 XX
 XX Cahoon EB, Kinney AJ, Klein TM, Lee J, Pearlstein RW;
 PI Rafalski JA, Shen JB, Thorpe CU, Tingey SV, Weng Z;
 XX
 DR WPI; 2003-567325/53.
 DR N-PSDB; AAL57712.
 XX
 PT New isolated polynucleotides encoding plant lipoxygenases, useful in
 PT genetic mapping, particularly in catalyzing hyperoxidation of
 XX polyunsaturated fatty acids.
 PS Cla:m 19; Page 23-25; 36pp; English.

AAU76413;
 08-MAY-2002 (first entry)
 Dimethylallyltryptophan synthase (DmaW) version #2.
 Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
 KW endophyte; forage; pasture; turf; land reclamation; soil conservation;
 KW ergot alkaloid; Neotyphodium; Acremonium; Epicchiole; Clavicipitaceae;
 KW Balansia; Balansiospis; Echinoschis; Atkinsonella; Myriogenospora;
 KW Neotyphodium; Paresichloe; knockout; antisense technology.
 XX
 OS Neotyphodium coenophialum.
 XX
 PN US6335188-B1.
 XX
 PD 01-JAN-2002.
 XX
 PF 03-MAR-2000; 2000US-00518657.
 XX
 PR 22-MAR-1999; 99US-0125490P.
 XX
 XX (KENT) UNIV KENTUCKY RES FOUND.
 PA
 PI Schardl CL, Wang J;
 XX
 DR WPI; 2002-163205/21.
 DR N-PSDB; ABK15521.
 XX
 XX New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 PT molecule) from fungi that are symbionts of commercially important
 PT grasses, useful to engineer ergot alkaloid-deficient symbionts.
 XX
 PS Claim 1; Col 15-18; 16pp; English.
 XX
 XX The invention describes an isolated dimethylallyltryptophan synthase
 CC (dmaW) nucleic acid (I). (I) is a dmaW sequence from epiphyte fungi
 CC Neotyphodium coenophialum symbiont of commercially significant grasses
 CC and is useful for expressing dmaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a dmaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epicchiole that lack dmaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of dmaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansia, Balansiospis,
 CC Echinoschis, Atkinsonella, Myriogenospora, Neotyphodium, and
 CC Paresichloe, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This is the amino acid sequence
 CC of dimethylallyltryptophan synthase (dmaW) described in the method of the
 XX invention
 XX
 SQ Sequence 450 AA;
 Query Match 39.0%; Score 904.5; DB 5; Length 450;
 Best Local Similarity 42.4%; Pred. No. 5.2e-78;
 Matches 180; Conservative 76; Mismatches 154; Indels 15; Gaps 6;
 7 AATLKPFFVVLQALNLSKDKTKWYSTAPMFATMAGAYDVHQAQYKFLCIHREVII 66
 4 AKTLHGEVYQTLSETDFDANNDRDLWHSHTAPNFQKILQTAIVSYIAQYQHSIYKSHII 63
 67 PALGPYPEKQPMHWKSHLTFGLPPELSFNYSKSLRFAFEPGLSLGTCKDDPNTQAI 126
 64 PFLGVYPTSRGE-RWLSILTRYCTPPELSLNCSDSIVRYTYEPIAATGSHLDPFNTFAI 122

XX This invention relates to novel nucleotide sequences which encode
CC proteins which have lipoxigenase activity. Lipoxigenases are membrane
CC bound ubiquitous enzymes which catalyse the hydroperoxidation of
CC polyunsaturated fatty acids in the first step of fatty acid metabolite
CC synthesis. Products of this pathway are found as signal molecules
CC involved in growth and development regulation. A knowledge of the amino
CC acid sequence of lipoxigenases may allow the understanding of plant
CC development and wound response. The polynucleotides, polypeptides and
CC lipoxigenases of the invention may therefore be useful in genetic mapping
CC and particularly for catalysing hydroperoxidation of polyunsaturated
CC fatty acids. The present sequence is the amino acid sequence of the
CC Balsam pear (Momordica charantia) lipoxigenase protein 2 of the invention
XX

XX SQ Sequence 880 AA;

Query Match 4.5%; Score 103.5; DB 6; Length 880;
Best Local Similarity 23.1%; Pred. No. 2.1;
Matches 96; Conservative 46; Mismatches 138; Indels 135; Gaps 24;

Qy 49 YDVHAQYKFLCIHR---EVIIPALG-----PYPEK---GQPMHWKSHLTRFGLPFELSFN 97
Db 221 YDV---YNDLCDPENGPNLVRPIILGGSDQYFPRGRGTGRPPARKOKHYESRLSDVMSLN 277

Qy 98 -YKSLRLFAPEPLGSLTGKDDPFTNQAIRPVLDLQKAMVPGLDLEWFDHFTKALVWSE 156
Db 278 IYVPRDENFGHLKWDPLGNLTKVLST-SIQPGLSEIFDSTPG-----EFDKP----- 324

Qy 157 BEARTLDRDIETP--VFKTQNKLAADLEPSGDIVLKYIYPRIKSIATGTPKERLMPDA 214
Db 325 KEVDLDLFRGFPIPLNIFK---NLTEDLAPP-----LTKA 356

Qy 215 IKAAD--KFGKVAIPLALE-----EFTAEAPTLLGHFLCLDLVKPSESRI 259
Db 357 FLRSDGRFLKYPFPQVQKNGKLGWRTDEFAREMIAGVNPPLII-----REL 403

Qy 260 KYCYMERQLDLASIEGIWTLNGRNDPET-----LDGL---DALRE-LWOLLPVTEGL 308
Db 404 EVFPPLSKLP-----HVVGNQNSTMTTEQIKHGLDGLTVDEAIKENKLYLDHHDAL 456

Qy 309 CPLPNCYEGCTSPQQLPIL---NFTLSP---KSALPEPQ-----IYFPAFGQN 353
Db 457 MPYLRRINSTKTYATRTLLFLKDDSTLKLPLAIELSLPHPOQDDEHGAISKLYFFAAGR- 515

Qy 354 DKTIAGLATFFSRGNGGLAKSPADLASYPDVP---LQTAHLQAWISPSYK 405
Db 516 -----VESAIW-QLAKEYAVVNDSGYHQLNSHLTHAVLEPFIITHR 558

RESULT 5
AAM50327
ID AAM50327 standard; protein; 521 AA.

XX AAM50327;
AC AAM50327;
XX
XX
DT 04-FEB-2002 (first entry)
XX
XX Human nucleotide binding site protein NBS-4.
DE
XX NBS-4; nucleotide binding site; human; antiinflammatory; antiapoptotic;
XX cytotatic; antimicrobial; anti-HIV; antiparkinsonian; antianemic;
XX neuroprotective; nootropic; cardiant; cerebroprotective; antiarthritic;
XX antidiabetic; immunosuppressive; thyromimetic; antibacterial;
XX tuberculostatic; virucide; signal transduction; vaccine; therapy;
XX diagnosis.
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH
FT Modified-site 11..14
FT Modified-site 23..26 /note= "Asn is N-glycosylated"

FT Modified-site 24..27 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 24..26 /note= "O-phosphorylated by protein kinase C"
FT Modified-site 39..42 /note= "Asn is N-glycosylated"
FT Domain 42..521 /label= Nucleotide-binding_site
FT Domain 47..62 /label= Kinase-1a_domain
FT Binding-site 52..59 /note= "P-loop"
FT Modified-site 55..60 /note= "ATP/GTP-binding site motif A (P loop)"
FT Modified-site 97..100 /note= "N-myristoylated"
FT Domain 116..132 /label= Kinase-2a_domain
FT Modified-site 118..121 /note= "Walker B Box"
FT Modified-site 138..141 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 142..145 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 146..151 /note= "N-myristoylated"
FT Domain 174..185 /label= Kinase-3a_domain
FT Modified-site 179..181 /note= "O-phosphorylated by protein kinase C"
FT Modified-site 204..207 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 220..223 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 232..235 /note= "Asn is N-glycosylated"
FT Modified-site 282..285 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 295..300 /note= "N-myristoylated"
FT Modified-site 312..315 /note= "Asn is N-glycosylated"
FT Modified-site 320..323 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 332..335 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 346..351 /note= "N-myristoylated"
FT Modified-site 355..358 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 377..379 /note= "O-phosphorylated by protein kinase C"
FT Modified-site 426..428 /note= "O-phosphorylated by protein kinase C"
FT Modified-site 462..465 /note= "O-phosphorylated by casein kinase II"
FT Modified-site 468..470 /note= "O-phosphorylated by protein kinase C"
XX WO200183753-A2.
XX 08-NOV-2001.
XX 03-MAY-2001; 2001WO-US014283.
XX 03-MAY-2000; 2000US-0201464P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Bertin J;
PI

DR WPI; 2002-041495/05.
DR N-PSDB; AA170682.
XX
XX Novel isolated polypeptide of nucleotide binding site useful as a vaccine
PT for preventing or treating diseases e.g. cancer, stroke, Alzheimer's
PT disease, Parkinson's disease, myocardial infarction, Crohn's disease.
XX
XX Claim 12; Fig 9A-C; 149pp; English.
XX
XX The present sequence is that of novel human NBS-4, as deduced from a
CC partial cDNA clone (see AA170682). The sequence is predicted not to be
CC full length. The human NBS-4 protein has a nucleotide binding site, which
CC is present in a number of proteins that transmit signals which activate
CC apoptotic and inflammatory pathways in response to stress and other
CC stimuli. The predicted full-length protein should also have a leucine-
CC rich repeat domain (LRR), which is also present in proteins involved in
CC apoptotic and inflammatory pathways, and a pyrin domain. The invention
CC provides NBS-2, NBS-3, NBS-4 and NBS-5 nucleic acids (see AA170680-85)
CC and polypeptides (see AA170682-90), as well as modulators of NBS-2, NBS-
CC 3, NBS-4 and NBS-5 activity or expression. These are expected to be
CC useful in the modulation of stress-related, apoptotic and inflammatory
CC responses. The nucleic acids and polypeptides are also expected to be
CC useful in diagnosis of such disorders and in screening assays used to
CC identify modulator compounds. The inflammatory and immune disorders
CC include, but are not limited to, chronic inflammatory diseases and
CC disorders, such as Crohn's disease, reactive arthritis, including Lyme
CC disease, insulin-dependent diabetes, organ-specific autoimmunity,
CC including multiple sclerosis, Hashimoto's thyroiditis and Grave's
CC disease, contact dermatitis, psoriasis, graft rejection, graft versus
CC host disease, sarcoidosis, atopic conditions, such as asthma and allergy,
CC including allergic rhinitis, gastrointestinal allergies, including food
CC allergies, eosinophilia, conjunctivitis, glomerular nephritis, and
CC infections including helminthic (e.g. leishmaniasis), viral (e.g. HIV),
CC and bacterial (e.g. tuberculosis and lepromatous leprosy). Apoptotic
CC pathway disorders include cancer, infectious diseases and autoimmune
CC disorders such as systemic lupus erythematosus, immune-mediated
CC glomerulonephritis and arthritis, as well as neurological diseases
CC including Alzheimer's disease, Parkinson's disease, amyotrophic lateral
CC sclerosis, retinitis pigmentosa and Huntington's disease, and
CC haematological diseases including aplastic anaemia, chronic neutropaenia
CC and myelodysplastic syndromes
XX
XX Query Match 4.4%; Score 102; DB 5; Length 521;
Best Local Similarity 22.3%; Pred. No. 1.4; Indels 94; Gaps 18;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
QY 79 MHWKSHL---TRFGLPFELSFNYSKLLRFAFEPGLSLGTGK-DDPFNTQAIRPVLQDLK 134
DB 67 LHWANGVLFOQRFSYFVYLSCHKIRYKETTFAELISLDWPDFDAPIEFMSQP--EKLL 124
QY 135 AMVPGDLEWDFHTKALVSVSEEEARTLLDRD-----IEIPVK-TQNKLAADLEPSGD 187
DB 125 FIIDGFE-----EIIISERSESDDGSPCTDWYQELPVTKILHSLKKEVLPLAT 175
QY 188 --IVLKTIVYIPRIKSI-----ATGTPKERMFDALKAADKFGKVATPLAILEE--- 233
DB 176 LLITIKTWFRDLKASLVNFCVQITGTDGLRVYFMRHFDSDSEVEKILQQLRKNETL 235
QY 234 FIAERAPTLGLHFLSC-----DLVKSERIKYVC-----MERQLDIA----- 271
DB 236 FHSCSAPMVCWTCVSCILKQPKVRYDIQSITQTTSLYAFNLSFSTAEVDLADDSWPG 295
QY 272 -----STEGITWTLNGRNDPET-LDGI-----DALRELWOLLFVTR-GLCFL----- 311
DB 296 QWRALCSLAEGLWSMNTTKEDTEIEGLEVFIDLSLVEFNILQKINDCGGCTTTHLS 355
QY 312 -----PNCFYBPGTSQSQLPFIIFNLTLSPKSALPEPQIVPPAFGQNDKTI 357
DB 356 QQBFAAAMSVLEEFREPPHSTKQPB-MXWMLQHVLLDKEAYWTFVPLFP-PGLANKMI 413
QY 358 ABGL 361

DB 414 AREL 417
XX
XX RESULT 6
AA017868
ID AA017868 standard; protein; 606 AA.
XX
XX AC AA017868;
XX
XX 20-AUG-2002 (first entry)
XX
XX Pyrin domain containing protein NALP13/Pyl7.
XX
XX Pyrin domain; PYD domain; antiinflammatory; antiparkinsonian;
KW antiarteriosclerotic; antipsoriatic; antibacterial; virucide;
KW neuroprotective; antiarthritic; antirheumatic; antiasthmatic;
KW neurotropic; osteopathic; nontropic; intracellular signal transduction;
KW inflammation; Alzheimer's disease; infection; psoriasis; asthma;
KW arteriosclerosis; multiple sclerosis; rheumatoid arthritis; sarcoidosis;
KW osteoarthritis; glomerulonephritis.
XX
XX Unidentified.
XX
XX WO200240668-A2.
XX
XX 23-MAY-2002.
XX
XX 30-OCT-2001; 2001WO-EP012545.
XX
XX 15-NOV-2000; 2000DE-01056687.
XX
XX 30-NOV-2000; 2000DE-01059595.
XX
XX (APOT-) APOTECH RES & DEV LTD.
XX
XX Tschopp J, Martinon F;
PI WPI; 2002-427093/45.
DR N-PSDB; AA147140.
XX
XX New DNA encoding protein with pyrin domain, useful for treating diseases
PT involving impaired signal transduction, particularly inflammation, also
PT proteins and antibodies.
XX
XX Claim 5; Fig 1; 116pp; German.
XX
XX The present invention relates the DNA and their encoded proteins, where
CC the proteins contain at least one PYD (pyrin) domain. These can be used
CC to treat diseases associated with impaired intracellular signal
CC transduction, particularly inflammation such as psoriasis,
CC arteriosclerosis, bacterial or viral infections (particularly meningitis
CC and pneumonia), multiple sclerosis, rheumatoid arthritis, asthma,
CC sarcoidosis, glomerulonephritis and osteoarthritis, and also Alzheimer's
CC and Parkinson's diseases. The present sequence is a protein of the
CC invention
XX
XX Query Match 4.4%; Score 102; DB 5; Length 606;
Best Local Similarity 22.3%; Pred. No. 1.7; Indels 94; Gaps 18;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
QY 79 MHWKSHL---TRFGLPFELSFNYSKLLRFAFEPGLSLGTGK-DDPFNTQAIRPVLQDLK 134
DB 232 LHWANGVLFOQRFSYFVYLSCHKIRYKETTFAELISLDWPDFDAPIEFMSQP--EKLL 289
QY 135 AMVPGDLEWDFHTKALVSVSEEEARTLLDRD-----IEIPVK-TQNKLAADLEPSGD 187
DB 290 FIIDGFE-----EIIISERSESDDGSPCTDWYQELPVTKILHSLKKEVLPLAT 340
QY 188 --IVLKTIVYIPRIKSI-----ATGTPKERMFDALKAADKFGKVATPLAILEE--- 233
DB 341 LLITIKTWFRDLKASLVNFCVQITGTDGLRVYFMRHFDSDSEVEKILQQLRKNETL 400

QY 234 FIABRAPTLGHFLSC-----DLVKPSBSRIKVC-----MERQDLIA----- 271
 Db 401 FHSCSAPMVCWTVCSCLPKQKRVYDLQSIQTQTTSILYAFFNSLFAEVDLADDSWPG 460
 QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWOLLPVE-GLCPL----- 311
 Db 461 QWRALCSLAIEGLSMNFTNKEDTIEGLEVFIDSLYEFNLQINDCGGTTTHLS 520
 QY 312 -----PNCFYEPGTSPOELPFIINFTLSPKSLPEPQIYFPAGQNDKI 357
 Db 521 FQEFFAAMSFLBEPREFPHSTKPOE-MQMLLQHVLLDKEAYWTPVWLF-PFGLLNKNI 578
 QY 358 AEG 361
 Db 579 AREL 582

RESULT 7
 ABO07113
 ID ABO07113 standard; protein; 635 AA.
 AC ABO07113;
 XX 13-AUG-2003 (first entry)
 XX Novel human protein NOV2a.
 XX NOVX; autoimmune disease; allergy; Alzheimer's disease; stroke;
 XX Parkinson's disease; Huntington's disease; multiple sclerosis; addiction;
 XX anxiety; pain; diabetes; glomerulonephritis; obesity;
 XX systemic lupus erythematosus; asthma; scleroderma; pancreatitis;
 XX graft versus host disease; ulcer; anaemia; cancer; trauma; infection;
 XX cardiomyopathy; atherosclerosis; hypertension; AIDS; Crohn's disease;
 XX acquired immunodeficiency syndrome; chromosomal mapping; tissue typing;
 XX forensic biology; predictive medicine; gene therapy; human.
 XX Homo sapiens.
 OS
 XX Best Local Similarity 22.3%; Pred. No. 1.8;
 XX Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
 XX 12-DEC-2002.
 PD
 XX 04-JUN-2002; 2002WO-US017558.
 PF
 XX 04-JUN-2001; 2001US-0295607P.
 PR 04-JUN-2001; 2001US-0295661P.
 PR 06-JUN-2001; 2001US-0296404P.
 PR 06-JUN-2001; 2001US-0296418P.
 PR 07-JUN-2001; 2001US-0296757P.
 PR 11-JUN-2001; 2001US-0297414P.
 PR 12-JUN-2001; 2001US-0297567P.
 PR 15-JUN-2001; 2001US-0298528P.
 PR 18-JUN-2001; 2001US-0299133P.
 PR 19-JUN-2001; 2001US-0299230P.
 PR 21-JUN-2001; 2001US-0299499P.
 PR 22-JUN-2001; 2001US-0300177P.
 PR 26-JUN-2001; 2001US-0300883P.
 PR 28-JUN-2001; 2001US-0301530P.
 PR 28-JUN-2001; 2001US-0301550P.
 PR 03-JUL-2001; 2001US-0302951P.
 PR 12-SEP-2001; 2001US-0318727P.
 PR 27-SEP-2001; 2001US-0325685P.
 PR 22-FEB-2002; 2002US-0358814P.
 PR 03-JUN-2002; 2002US-00161927.
 XX (CURA-) CURAGEN CORP.
 XX Zerhusen BD, Kekuda R, Spytek KA, Shenoy SG, Miller CE, Hjalt T;
 XX Gerlach VL, Baumsartner JC, Guo X, Gangolli EA, Vernet CAM;
 XX Padigar M, Li L, Pena CEA, Gorman L, Anderson DW, Edinger SR;
 XX Paturajan M, Stone DJ;

WPI; 2003-140585/13.
 N-PSDB; ACD13187.
 Novel isolated NOVX polypeptide useful treating or preventing disorders
 or syndromes such as autoimmune disease, allergies, Alzheimer's disease,
 stroke, Parkinson's disease, Huntington's disease or multiple sclerosis.
 Claim 1; Page 87; 408pp; English.
 The invention describes an isolated NOVX polypeptide (I) comprising a
 sequence selected from a sequence (S1) of 1121, 635, 299, 1720, 176, 583,
 214, 395, 1098, 134, 427, 1333, 407, 806, 804, 1253, 382, 1045, 284, 496,
 506, 759, 390, 133, 215, 240, 1069, 116, 439, 1138, 477, 316, 269, 219,
 305, 406, 460, 365, 380, 829 or 326 amino acids fully defined in the
 specification, and the mature form of Si. (I) is useful for treating or
 preventing a pathology associated with (I) in a subject, preferably
 human, or for identifying an agent that binds to (I), where the agent is
 a cellular receptor or a downstream effector. (I), a polynucleotide (II)
 encoding (I) or an anti-(I)-antibody (V) is useful treating or preventing
 disorders or syndromes such as autoimmune disease, allergies, Alzheimer's
 disease, stroke, Parkinson's disease, Huntington's disease, multiple
 sclerosis, addiction, anxiety, pain, diabetes, glomerulonephritis,
 systemic lupus erythematosus, asthma, scleroderma, graft versus host
 disease, pancreatitis, obesity, ulcers, anaemia, cancer, trauma, viral,
 bacterial or parasitic infections, cardiomyopathy, atherosclerosis,
 hypertension, acquired immunodeficiency syndrome (AIDS) or Crohn's
 disease. (I), (II) or (V) is useful in screening assays, detection assays
 (e.g., chromosomal mapping, tissue typing, forensic biology), predictive
 medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical
 trials and pharmacogenomic), and in methods of treatment (e.g.,
 therapeutic and prophylactic). (II) is useful in gene therapy, to express
 (I), to detect NOVX mRNA or a genetic lesion in a NOVX gene, and to
 modulate NOVX activity. This is the amino acid sequence of a novel human
 NOV protein

Sequence 635 AA;
 Query Match 4.4%; Score 102; DB 6; Length 635;
 Best Local Similarity 22.3%; Pred. No. 1.8;
 Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
 QY 79 MHWKSHL---TRGLPELSPFNYSKSLRPAFPLGSLTGK-DDPNTCAIRPVLDLX 134
 Db 67 LHWANGVLFQCRFSYVFLSKIRYKETTFAELISLDWDFDAPIEFMSQF--EKLL 124
 QY 135 AMVEGLDLEWDFHTKALVSEEARLDRD-----IEIPVEK-TQNKLAADLPSPGD 187
 Db 125 FIIDGFE-----EIIISSESSSLDDGSPTDWMYQELPVTKILHSLKXELPLAT 175
 QY 188 --IVLKYIYPRKSI-----AGTPKPERLWFAKAADKFGKVPALIEB---- 233
 Db 176 LLITIKTFVFRDLKASLWNPFCVQITGTGDDLRVYFMRFHFDSSSEVEKILQLRKNETL 235
 QY 234 FIABRAPTLGHFLSC-----DLVKPSBSRIKVC-----MERQDLIA----- 271
 Db 236 FHSCSAPMVCWTVCSCLPKQKRVYDLQSIQTQTTSILYAFFNSLFAEVDLADDSWPG 295
 QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWOLLPVE-GLCPL----- 311
 Db 296 QWRALCSLAIEGLSMNFTNKEDTIEGLEVFIDSLYEFNLQINDCGGTTTHLS 355
 QY 312 -----PNCFYEPGTSPOELPFIINFTLSPKSLPEPQIYFPAGQNDKI 357
 Db 356 FQEFFAAMSFLBEPREFPHSTKPOE-MQMLLQHVLLDKEAYWTPVWLF-PFGLLNKNI 413
 QY 358 AEG 361
 Db 414 AREL 417
 RESULT 8
 AAB11500

XX AAB11500;
AC
XX
DT 13-MAR-2001 (first entry)
XX
DE C. sativus LOX protein.
XX
KW LOX; lipoxigenase; cucumber; plant; 6-hydroperoxy-gamma-linolenic acid;
KW 9-hydroperoxy-gamma-linolenic acid; 6,9-hydroperoxy-gamma-linolenic acid;
KW gamma-linolenic acid.
XX
OS Cucumis sativus.
XX
PN DE19914464-A1.
XX
PD 05-OCT-2000.
XX
PF 30-MAR-1999; 99DE-01014464.
XX
PR 30-MAR-1999; 99DE-01014464.
XX
PA (IPBP-) IPB INST PFLANZENBIOCHEMIE.
XX
PI Feussner I, Hornung E;
XX
XX WPI; 2001-103874/12.
DR
XX Preparation of plant lipoxigenase with altered position specificity
PT comprises replacing at least one amino acid in the wild type lipoxigenase
PT amino acid sequence.
XX
PS Disclosure; Fig 5; 14pp; German.
XX
CC This invention describes a novel preparation of plant lipoxigenase (I)
CC with altered position specificity which comprises replacing at least 1
CC amino acid in the wild type lipoxigenase amino acid sequence. The method
CC also describes (1) (I) prepared by the method above; (2) nucleic acid
CC encoding (I); (3) a vector comprising the nucleic acid of (2); (4) a cell
CC comprising the vector of (3); (5) a plant or plant part comprising a host
CC cell as in (4); (6) preparation of 6-, 9- and/or 6,9-hydroperoxy-gamma-
CC linolenic acid comprising reacting gamma-linolenic acid with (I); and (7)
CC a gamma-linolenic acid derivative comprising a hydroperoxy group or a
CC hydroxy group at position 6
XX
SQ Sequence 878 AA;

Query Match 4.3%; Score 101; DB 4; Length 878;
Best Local Similarity 21.7%; Pred. No. 3.7;
Matches 94; Conservative 49; Mismatches 129; Indels 162; Gaps 26;
QY 61 HREVIIPALG-----PYPEK---GQPMHWKSHLTRFGLPELSFNYSKSLRFAPEPLGS 112
DB 235 HR-----PILGGTTEYPYPRRGRTGRPSRDH-----NYESRL-----SPIMS 273
QY 113 LT--GTKDDPFN-----TQAIRPVLDLKAMVPGLDLEWFDHFTKALVVS 155
DB 274 LDIVPKDENFGHLKXSDFLGYTLKALSISIKPGLQSFVTP-----NEFDNF----- 322
QY 156 EBEARTLLDRDIEIP--VFQTKNKLADLEPSGDIVLKYIYPRIKSIATGTPKERLMPD 213
DB 323 -KEVDNLFERGFPIPFNAFKT---LTEDLTP-----PLFKALVRNDGKFLKFP 367
QY 214 AIKAADKFGKV--ATPLAILEEFTAEAPTLLGHFLSCDLVKPSSESRIKYVCMERQLDLA 271
DB 368 TPVV-KONKIGWSTDEFAREMLAGNPLLI-----RLEAFPTSKLDP- 412
QY 272 SIEGIWTNGRNDPDT-----LDGLDALRELWQ-----LLPVTEGLCPNCFVEPGT 320
DB 413 -----NVYGNQNSTIEEHKGLDGLTVDEAMQNRLYIVDFHDALMPLYLRMNTATST 466
QY 321 SPQQLPFI--NFTLSP--KSALEPQO-----LYEFAFGQNDKTIAEGLATFF 365
DB 467 KTVATRRLLLKDDGTGLKLVIELALPHPOQDQGLAISKLKYFFAENGUVQKSI----- 518

QY 366 ESRGWGLAKSYPADLASYYPDVLDQTANHLQAWISFSYKGGKPKYMSVYLHT---FEAFS 422
DB 519 ---W-QLAKAY-----VTVDVGYHOLISHWLTHAVLEPFV 551
QY 423 AAA-QEVAMCHDGH 435
DB 552 IATHRQLSVLHPH 565
RESULT 9
AAB86036
ID AAB86036 standard; protein; 878 AA.
XX
AC AAB86036;
XX
XX 13-JUL-2001 (first entry)
XX
DE Cucurbit LBLOX protein SEQ ID 4.
XX
KW Cucurbit; LBLOX; fatty acid metabolism; lipid metabolism;
KW plant oil-production; transgenic plant.
XX
OS Cucumis sativus.
XX
PN DE19950921-A1.
XX
PD 26-APR-2001.
XX
PF 21-OCT-1999; 99DE-01050921.
XX
PR 21-OCT-1999; 99DE-01050921.
XX
PA (BADI) BASF AG.
XX
PI Kindl H, May C, Feussner I;
XX
DR WPI; 2001-274658/29.
DR N-PSDB; AAF88022.
XX
XX New isolated nucleic acid encoding sequence that targets proteins to
PT lipid bodies, useful for producing transgenic plants for lipid and fatty
PT acid production.
XX
PS Disclosure; Page 18-21; 30pp; German.
XX
CC This invention describes a novel isolated nucleic acid sequence (I),
CC encoding a polypeptide, comprising a sequence (Ia) involved in fatty acid
CC or lipid metabolism, and a targeting sequence (Ib). (I) are used to
CC produce oil-producing transgenic plants or eukaryotic microorganisms, for
CC production of lipids or derived fatty acids. This sequence represents a
CC Cucumis sativus (cucumber) LBLOX protein which is described in the method
CC of the invention
XX
SQ Sequence 878 AA;

Query Match 4.3%; Score 101; DB 4; Length 878;
Best Local Similarity 21.7%; Pred. No. 3.7;
Matches 94; Conservative 49; Mismatches 129; Indels 162; Gaps 26;
QY 61 HREVIIPALG-----PYPEK---GQPMHWKSHLTRFGLPELSFNYSKSLRFAPEPLGS 112
DB 235 HR-----PILGGTTEYPYPRRGRTGRPSRDH-----NYESRL-----SPIMS 273
QY 113 LT--GTKDDPFN-----TQAIRPVLDLKAMVPGLDLEWFDHFTKALVVS 155
DB 274 LDIVPKDENFGHLKXSDFLGYTLKALSISIKPGLQSFVTP-----NEFDNF----- 322
QY 156 EBEARTLLDRDIEIP--VFQTKNKLADLEPSGDIVLKYIYPRIKSIATGTPKERLMPD 213
DB 323 -KEVDNLFERGFPIPFNAFKT---LTEDLTP-----PLFKALVRNDGKFLKFP 367
QY 214 AIKAADKFGKV--ATPLAILEEFTAEAPTLLGHFLSCDLVKPSSESRIKYVCMERQLDLA 271

368 TPVV-KDKGWTDEFAEMLAGNPLLI-----RLEAFPTSKLD? 412
 272 SIEGWTNGRNDPET-----LDGLDALRELWQ-----LLPVTEGLCLPNCFYEPGT 320
 413 -----NVYQNGSTIEEHIKHLGLTVDEAMKQNLIVDFHDLMPYLFRMNTATST 466
 321 SPQQLPFI---NFTLSP---KSLPEPO-----IYFPAFGQNDKTIAGELATPF 365
 467 KYATRTLLKDDGTGLKPLVIELPHPODQGLAISKLYFAENGVSQSI----- 518
 366 ESRGWGLAKSYPADLASYPVDVLOTANHLQAWISFYKPKPMYSVLIHT---PEAFS 422
 519 ----W-QLAKAY-----VTWVDVGYHQLISHLHTHAVLEPFV 551
 423 AAA-QEVAMCHDGH 435
 552 IATHRQLSVLPHIH 565

RESULT 10
 ADE54304
 ID ADE54304 standard; protein; 503 AA.
 XX
 AC ADE54304;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein BAA20354, SEQ ID NO 107.
 XX
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.
 XX
 PN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GENO) GEN HOSPITAL CORP.
 PA (FAR) BAYER AG.
 XX
 XX Woolf C, D'urso D, Befort K, Costigan M;
 XX
 XX WPI; 2003-268312/26.
 DR GENBANK; BAA20354.
 DR

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.
 Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the

CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 503 AA;

Query Match 4.3%; Score 99.5; DB 7; Length 503;
 Best Local Similarity 19.3%; Pred. No. 2.3;
 Matches 92; Conservative 62; Mismatches 190; Indels 133; Gaps 19;

QY 39 MFATMMAGAGYDVHAQYKF-LCIHREVIIPALGPYPEK-GQPMHWKSHLTFEG-LPEELS 95
 Db 26 LLSTILLIACFTSLVYLFLAVGHVQLPAGAKSPPIYSPFPLGHATAFGKSPIEFL 85
 QY 96 FN-YSKSLRLRFAPEPLGS-----LTKDPPFNTOAI-----RPV----- 129
 Db 86 ENAYEKYGPVPSFTMVGKTFYLLGSDAALLFNKKNEDLNAERVYGRLLTPVPFGKVAY 145
 QY 130 -----LQDLKANVPGLDLEWFDHFTKALVSEEAR----- 160
 Db 146 DVPNAVLEQKKILKSGLNIA---HFQYVSIIEKEAKYFKSWGESGERNVFEALSELI 202
 QY 161 -----TLLDRDIBIPVFKTKNLAADLE-----PS-----GDIVLKT 192
 Db 203 ILTASHCLHGKEIRSQLNEKVAQYADLDGFSHAAWLLFGWLPSPFRDRRAHREIKN 262
 QY 193 YIYPRI-KSIATGTPKERMFDALKAADKFGKVATP-----LAILBEPTAERAPTL 242
 Db 263 IFYKAIQKRRLSKPEADILQTLDDSTYKDGREPLTDDEIAGMLIGLLLAGQHTSSTTSAW 322
 QY 243 LGHFLSCDLVKPSESRI---KVVC-----MERQLDLASIE-----GIWTL 279
 Db 323 MGFFLARD--KPLQDKCYLEQKTVCGEDLPPLYEQKDLNLDRCIKETLRLRPPIWTM 380
 QY 280 NGRNDPETLDGLDALRELWQLLPVTEGLCPNCFYEPGTSPQEQLPFIINFTLSPKSA 339
 Db 381 MRMAKTPQTIVAG-----YTTPPGHQVCVSPVTNQRKDSWVERLDENPDYQLQNP 432
 QY 340 LPEQIYPPAFGQNDKTIAGELATPFESRCGWSGLAKSYPADLAS-YVPDVLQTAH 395
 Db 433 SGEKFAVTPFGAGRHRCIGENFAVVKTIWSTMRLRYEFDLNGIPPSVNYTTLIH 489

RESULT 11
 ADE54292
 ID ADE54292 standard; protein; 503 AA.
 XX
 AC ADE54292;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein BAA20354, SEQ ID NO 95.
 XX
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.
 XX
 PN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX

26	Db	26	LLSTLLIACFTLSVLVFLFLAVGHVQLPAGAKSPPIYISPTFLGHATAFGKSPIEFL	85
96	Qy	96	FN-YSKSLRLFAPEPLGS-----LTGTKDDPFNTQAI-----RPV-----	129
86	Db	86	ENAYEKGVFSTWVGKTTFTYLLGSDAAALLFNSKNEDLNABEVYGRLTTFVFGKWAY	145
130	Qy	130	-----LQDLKAWPGDLDFHTKVLVVBEEAR-----	160
146	Db	146	DVNAVFELEQKTLKSLNIA--HFQKYYSIIIEKEAYFKSWGSEGRNVFEALSELI	202
161	Qy	161	-----TLLDRDIEI PVFKTCNKLAADLE-----PS-----GDIVLKT	192
203	Db	203	ILTASHCLHGKEIRSQLNEKVAQYADLDGCFSHAAWLLPGWLPLPSRRDRRAHREIKN	262
193	Qy	193	YIYPRI-KSIATGTPKPERLMFDAIKADKFGKVATP-----LAILIEFIAERAPT	242
263	Db	263	IFYKAIQKRRLSKEPABDIIQTLTLDSTYKQGRPLTDEIAGMLIGLLLAGOHTSSTISAW	322
243	Qy	243	LGHFLSCDLVKPSESRI---KYVC-----MERQLDASIF-----GIWTL	279
323	Db	323	MGFFLARD--KPLQDKCYLQKTVCGEDLPLPTVEQLKOLNLLDRICIKETLRLRPPIMTW	380
280	Qy	280	NGRNRNPETLDGLDALRELWOLLI PVTEGLCPLPNCVFYEPGTSPOEQLPFIINTFLSPKSA	339
381	Db	381	MRMAKTPQTWAG-----YTPPGHQVCVSPVNRQLKDSWVERLDFNDRYLQDNPA	432
340	Qy	340	LPPEQIYFPFAGQNDKTIAGSLATFFESRSGWGLAKSPADLAS-YVPDVLQDTANH	395
433	Db	433	SGSKFAVYPPGAGRHCIGENFAYVQIKTWTSTMLRYEEDLINGFYSVNYTTMIIH	489

RESULT 15

RESOL 13	RESOL 13
ADE54300	ADE54300
ID	ADE54300 standard; protein; 503 AA.

AA Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
KW

14-AUG-2002: 2002WO-US025765.

AA
PR 14-AUG-2001; 2001US-0312147P.XX
PA (GEHO) GEN HOSPITAL CORP.XX
PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat

RESULT 14
ADE62266
ID ADE62266 standard; protein; 503 AA.

Rat Protein Q64654, SEQ ID NO 8195.

Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury; chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

OS *Rattus norvegicus*.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.
vv

PR 14-AUG-2001; 2001US-0312147P.
PB 01-NOV-2001: 2001US-0346382D

PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNR)) in an animal (e.g. gene therapy). The sequence presented is a rat protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at [ftp.wipo.int/pub/published](http://wipo.int/pub/published) pct sequences.

SQ Sequence 503 AA;

Query Match 4.3%; Score 99.5; DB 7; Length 503;

Matches 92; Conservative 62; Mismatches 190; Indels 133; Gaps 19;

Run on: July 22, 2004, 18:44:52 ; Search time 19 Seconds
(without alignments)
1187.397 Million cell updates/sec

Sequence: 1 MEISKKAATLLPKPFVLSQ.....FEAFSAAQEVAMCHDGNP 437

ues

meters: 389414

A_COMB.per:*

B_COMB.pep:*

A-COMB pep: *

CTUS COMB. pep: *

ackfiles1.pep: *

edicted by chance to have a
ore of the result being printed,
al score distribution.

53E

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2332	100.0	437	4	US-09-472-364-2	Sequence 2, Appli
2	907.5	39.1	448	4	US-09-518-650-2	Sequence 2, Appli
3	904.5	39.0	450	4	US-09-518-657-4	Sequence 4, Appli
4	94.5	4.1	430	4	US-09-198-452A-701	Sequence 701, App
5	93	4.0	630	4	US-09-489-039A-13023	Sequence 13023, A
6	92	4.0	833	2	US-08-844-086-2	Sequence 2, Appli
7	92	4.0	833	3	US-09-018-211-2	Sequence 2, Appli
8	90.5	3.9	835	4	US-09-758-282B-155	Sequence 155, App
9	90.5	3.9	835	4	US-09-758-282B-343	Sequence 243, App
10	90	3.9	4302	3	US-08-658-136-5	Sequence 5, Appli
11	90	3.9	4302	4	US-09-052-463-8	Sequence 8, Appli
12	90	3.9	4302	4	US-08-422-583-8	Sequence 8, Appli
13	90	3.9	4302	4	US-09-052-262-8	Sequence 8, Appli
14	90	3.9	4303	2	US-08-460-751-2	Sequence 2, Appli
15	90	3.9	4339	4	US-09-052-469-6	Sequence 6, Appli
16	90	3.9	4339	4	US-08-422-583-6	Sequence 6, Appli
17	90	3.9	4339	4	US-09-052-263-6	Sequence 6, Appli
18	89	3.8	1317	3	US-09-083-521-7	Sequence 7, Appli
19	88.5	3.8	1420	4	US-09-125-635-4	Sequence 4, Appli
20	88	3.8	660	4	US-09-268-347-45	Sequence 45, Appli
21	88	3.8	1745	2	US-09-031-485-33	Sequence 33, Appl
22	88	3.8	1745	2	US-08-847-429A-33	Sequence 33, Appl
23	88	3.8	1745	4	US-09-065-474-33	Sequence 33, Appl
24	88	3.8	1745	4	US-09-557-034-33	Sequence 33, Appl
25	87.5	3.8	322	3	US-08-469-318-138	Sequence 138, App
26	87.5	3.8	322	3	US-08-469-318-154	Sequence 154, App
27	87.5	3.8	322	3	US-08-468-609A-138	Sequence 138, App

Db 181 DLEPFGDVLKTYIYPRISATGTPKRLMFDALKAADKFGKVAATPLALEEFAERAP 240
Qy 241 TLGHFLSCDLVKPSSRIKVCYMERQDLASTEGITWINGRNDPDTLDGLDALRELWQ 300
Db 241 TLGHFLSCDLVKPSSRIKVCYMERQDLASTEGITWINGRNDPDTLDGLDALRELWQ 300
Qy 301 LLPVTEGLCPNCFYEPGTSPOQLPFIINFTLSPKSLPPEQIYFPAGQNDKTIABG 360
Db 301 LLPVTEGLCPNCFYEPGTSPOQLPFIINFTLSPKSLPPEQIYFPAGQNDKTIABG 360
Qy 361 LATFFESRGWGLAKSYPADLASYPDVLQTAHILQAMISPSYKGGKPYMSVYLHTFEA 420
Db 361 LATFFESRGWGLAKSYPADLASYPDVLQTAHILQAMISPSYKGGKPYMSVYLHTFEA 420
Qy 421 FSAAAQEVAMCHDGNP 437
Db 421 FSAAAQEVAMCHDGNP 437

RESULT 2
US-09-518-657-2
; Sequence 2, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; APPLICANT: Wang, Jinghong
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; FILE OF INVENTION: Which Encode Therefor and Related Methods
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; EARLIER APPLICATION NUMBER: 60/125,490
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Neotyphodium coenophialum
US-09-518-657-2

Query Match 39.1%; Score 907.5; DB 4; Length 448;
Best Local Similarity 41.3%; Pred. No. 1.5e-86;
Matches 184; Conservative 78; Mismatches 157; Indels 27; Gaps 7;

Qy 7 AATLLPKPFVYLSQLNLNKHDTKMYSTAPMTAGAGYDVHQAQYKFLCIHREVII 66
Db 4 AKTLHQEVYTLSETDFANNORLWHSAPMFKQLQTANYSDAQRHGLIYKSHVI 63
Qy 67 PALGPYPERQPMHWKSHLTRFGLPFELSFNYSKSLRFAPEPLGSLGTGKDDPNTQAI 126
Db 64 PFLGVYPTRSGE-RWLSILTRYGTPPELSLNCSDSVRYTYEPINAATGSHLDPNTFAI 122
Qy 127 RVLQDLKAMVPGLDLEWFDHFTKALVVEEBEARTLLDRDIEIPVFKTONKLAADLEPSG 186
Db 123 WEALKKHISQPGIDLEWFSYFQELTLDANESTYLHSONLVKEQIKTONKALDLK--G 180
Qy 187 D-IVLKYIYPRIKSIATGTPKRLMFDAL-KAADKFGKVAATPLALEEFAERAP--- 240
Db 181 DKFVLKTYIYPELKSATGKSVQELVFGSVRKLAKQHKSIKIRPAFEMLEDYVQSRNKFTT 240
Qy 241 -----TLL-CHFLSCDLVKPSSRIKVCYMERQDLASTEGITWINGRNDPDTLDGLDA 294
Db 241 DSHNTLLSRLSLSCDLISPTKSRKIYLLERMSVSLPAMEDLWTLGGRREDQSTIEGLEM 300
Qy 295 LRELWOLLPVTEGLCPNCFYEPGTSPOQLPFIINFTLSPKSLPPEQIYFPAGQND 354
Db 301 IRELWGLLNSPGLRAYPEPPLGALPNEQLPSMANYTLHNDPIPEQVYFTVFGMD 360
Qy 355 KTAEGLATFFESRGWGLAKSYPADLASYPDVLQTAHILQAMISPSYKGGKPYMSVY 414
Db 414 KTAEGLATFFESRGWGLAKSYPADLASYPDVLQTAHILQAMISPSYKGGKPYMSVY 414

Qy 415 LHTEFE-----AFSAAAQEV 428
Db 421 LHSFETGKWPVPEGLIAFDACRRDL 446

RESULT 3
US-09-518-657-4
; Sequence 4, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; APPLICANT: Wang, Jinghong
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; FILE OF INVENTION: Which Encode Therefor and Related Methods
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; EARLIER APPLICATION NUMBER: 60/125,490
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Neotyphodium coenophialum
US-09-518-657-4

Query Match 39.0%; Score 904.5; DB 4; Length 450;
Best Local Similarity 42.4%; Pred. No. 3.2e-86;
Matches 180; Conservative 76; Mismatches 154; Indels 15; Gaps 6;

Qy 7 AATLLPKPFVYLSQLNLNKHDTKMYSTAPMTAGAGYDVHQAQYKFLCIHREVII 66
Db 4 AKTLHQEVYTLSETDFANNORLWHSAPMFKQLQTANYSDAQRHGLIYKSHII 63
Qy 67 PALGPYPERQPMHWKSHLTRFGLPFELSFNYSKSLRFAPEPLGSLGTGKDDPNTQAI 126
Db 64 PFLGVYPTRSGE-RWLSILTRYGTPPELSLNCSDSVRYTYEPINAATGSHLDPNTFAI 122
Qy 127 RVLQDLKAMVPGLDLEWFDHFTKALVVEEBEARTLLDRDIEIPVFKTONKLAADLEPSG 186
Db 123 WEALKKHISQPGIDLEWFSYFQELTLDANESTYLHSONLVKEQIKTONKALDLK--G 180
Qy 187 D-IVLKYIYPRIKSIATGTPKRLMFDAL-KAADKFGKVAATPLALEEFAER--APT- 241
Db 181 DKFVLKTYIYPELKSATGKSVQELVFGSVRKLAKQHKSIKIRPAFEMLEDYVQSRNKVPT 240
Qy 242 -----LLGHFLSCDLVKPSSRIKVCYMERQDLASTEGITWINGRNDPDTLDGLDA 294
Db 241 DSENTPLSRLSLSCDLVSPKSRKIYLLERMSVSLPAMEDLWTLGGRREDQSTIEGLEM 300
Qy 295 LRELWOLLPVTEGLCPNCFYEPGTSPOQLPFIINFTLSPKSLPPEQIYFPAGQND 354
Db 301 IRELWGLLNSPGLRAYPEPPLGALPNEQLPSMANYTLHNDPIPEQVYFTVFGMD 360
Qy 355 KTAEGLATFFESRGWGLAKSYPADLASYPDVLQTAHILQAMISPSYKGGKPYMSVY 414
Db 361 MEVTNALTKEFMREHSDMSKACLESFPHHNEALNYIHSYISFSYRNKNKPYLSVY 420
Qy 415 LHTEFE 419
Db 421 LHSFE 425

RESULT 4
US-09-198-452A-701
; Sequence 701, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 701
LENGTH: 430
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-701

Query Match
Best Local Similarity 4.1%; Score 94.5; DB 4; Length 430;
Matches 64; Conservative 41; Mismatches 126; Indels 51; Gaps 15;

QY 25 SNKHDKWYSAPFATMAGAG-YDVHAQYKFLCI-----HREVIIPALGYPPEKG 76
DB 104 SHRSNKKFYI-ILPFRYERQAGRYQHOFQVEAIGVRHPLRDAEVALLLWDFYSRVG 162
QY 77 QPMHNSHLTRGLPELSFNYSKSLRFAPEPLGSLTGKDDPNTQAIR-----PVL 130
DB 163 L-CHMOIQNLFLG-GSETFRDKVLRAYLKESMGELSALSQCRFTNVLRLDSKEPED 220
QY 131 QDLKAMVPGI-----DLEWFDHFTKALVSEEEARTLLDRIDRIPVFKTONKLAADLE 183
DB 221 QEIIRQAPILDVSDLDKYNELDAL-----RVLEBP-YAINPRLVRGLD 267
QY 184 PGSDIVLK-TYIYPRKSIATGTPKRLMFDKAADKFGKATPLAILEEFIAERATL 242
DB 268 YYSDLVFEATTTFOEVSVALGGGR---YDGLISA--FGASLPACGFGVGLERAIQTL 321
QY 243 LGHFLSCDLVKPS-ESRIKVCYMERQDLASIEGWTLLNGRR 283
DB 322 LAQ-----KRIEQPHKLRLPNEPDADQFCLE--WSQHLRR 357

RESULT 5
US-09-489-039A-13023
Sequence 13023, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: 2709.2004001
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 13023
LENGTH: 630
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13023

Query Match
Best Local Similarity 4.0%; Score 93; DB 4; Length 630;
Matches 56; Conservative 25; Mismatches 64; Indels 110; Gaps 14;

QY 69 LGYPPEKGMHWK-----SHLTFGLPELSFNYSKSLRF--APEPLGSLTGKDDPF 121
DB 320 LQRPDLFQGMDELERVVRHLPFLHATYDESIRMLDVPEKV-----NRDIEF 374
QY 122 NTOAIRPVLQDLKAMVGLDLEW-FDHFTKALVSEEEARTLLDRDIEIPVFKTONKLA 180
DB 375 N-----GLHWFFDH-----AETITERNIE----- 393
QY 181 DLEPSGDIVLTKTYIPRIKSIATGTPKRLMFDKAADKFGKAT-----PLAILEEF 235
DB 394 -----RVKALGGGIAVQHRMAFQGGFYFVDRYGKEAVGHTPPVA---KML 434
QY 236 AERAPTLGHFLSCDLVKPSERIKVCYMERQDLASIEGWTLLNGRRNDPDTLGL--- 292

DB 435 ALDVPVGLG-----TDATRVASNPWTAL-----YWLVSGR-----TVGGWAMY 473
QY 293 -----DALRELM 299
DB 474 DDANRLPRDVALELM 488

RESULT 6
US-08-844-086-2
Sequence 2, Application US/08844086
Patent No. 5866390
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5866390el Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,086
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-844-086-2

Query Match
Best Local Similarity 4.0%; Score 92; DB 2; Length 833;
Matches 96; Conservative 65; Mismatches 192; Indels 140; Gaps 25;

QY 16 YVLSQALNLSNKHDKW--WYSTAPMFATMAGAGYDVHAQYKFL-----CIHREVIIP- 67
DB 119 YDWDREVNTTDPNYKWTQWTF-----KLYEKLAYAEVPMVVEELGTATANEVLPD 174
QY 68 --ALGYPPEKGMHWKSHLTFGL--LPPPEL-SFNYSKSLRFAPEPLGSLGATK--- 117
DB 175 GTSERGGYVVRKPMRQWMLKITYAESLLNDLDELDSSESIKDMQRNWKSTGANVTF 234
QY 118 -----DDFNTQAIRPVL---QDLKAMVGLDLEWFDHFTKALVSEEEARTLLDRDIEI 169
DB 235 KVKGTDEFTVFTTRPDTLFGATFTVLAP-----EHELVDATISSEQAFAVAD--- 282
QY 170 PVFKTONKLAADLPSGDIVLTKTYIPRIKSIATGTPKRLMFDKAADKFGKATPLA 229
DB 283 --YKHQASLKDLARTDLAKETGWYTGAYAINPNVNGKEMPIWIADYVLSYGTGAV--- 337
QY 230 ILEEFIAERAPTLGHFLSCDLVKPSERIKVCYMERQDLASI-----RGIWTLN 280

Db 338 -----MAVPAHQ-RDWEFAKQFDLPVVELEGVVEAAAYTED 375
Qy 281 GRRNDPETHLGL-----DALRELWQL-----PVTEGLC-----PLNCFYE 317
Db 376 GLHVSDFDLGKEDAIKIVACLEKGGQKVTYRLDWLFSRQRYWGEPIPIHWE 435
Qy 318 PGTS---PQSLPFIINFTLSPKSALPEQIYPPAFQND---KTIAGLATFFESR--- 368
Db 436 DGTSTAVPETELPLVPVT---KDIRPSGTGESPLANLTDWLEVTREDGVKGRRETNTMP 492
Qy 369 GWGLAKSY-----PADLASYPVDVLOTAN-----HL---QAWISFSY-- 404
Db 493 QWAGSSWYLYRYIDPHNTEKLADEDLKQWLPVDIVVGGAEHVLHLLYARFWHKLFDL 552
Qy 405 ---KGKPYMSVY 414
Db 553 GWVTEKPEFQKLF 565

RESULT 7

US-09-018-211-2
; Sequence 2, Application US/09018211
; Patent No. 6048716
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 6048716el Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,211
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 06/844,086
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: 9607993.4
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31457-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 833 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-018-211-2

Query Match 4.0%; Score 92; DB 3; Length 833;

Best Local Similarity 19.5%; Pred. No. 2,2;
Matches 96; Conservative 65; Mismatches 192; Indels 140; Gaps 25;

Qy 68 ---ALGPYPEKQPM-HWKSHLTRFG--LPFEL-SFNYSKSLRFAFEPLGSLTGTK--- 117
Db 175 GTSERGYPVVKPMRQWMLKITAYAERLNDLDELQWSESIKDMQSNWIGKSTGANVTF 234
Qy 118 -----DDPNTQAIRPVL---QDLKAMVPGLDLEWFDHFTKALVVSEEBARTLLDRDIEI 169
Db 235 KVGTDKEFTVFTTRPDTLFGATFTVLAP-----EHELVDAITSEQAEAVAD----- 282
Qy 170 PVFTQNKLAADLPESGDIVLKVITYPRIKSIATGTPKRLMFDKAKADKFGKATPLA 229
Db 283 --YKHQASLKSDDLARTDLAKEKTGWGTGAYAINPVNGENKPIWIADVVLASYGTGAV--- 337
Qy 230 ILEPIAERAPTLGHLFSLCDLVKPSRSRIKVMCMERQDLASIT-----EGIWTLN 280
Db 338 -----MAVPAHQ-RDWEFAKQFDLPVVELEGVVEAAAYTED 375
Qy 281 GRRNDPETHLGL-----DALRELWQL-----PVTEGLC-----PLNCFYE 317
Db 376 GLHVSDFDLGKEDAIKIVACLEKGGQKVTYRLDWLFSRQRYWGEPIPIHWE 435
Qy 318 PGTS---PQSLPFIINFTLSPKSALPEQIYPPAFQND---KTIAGLATFFESR--- 368
Db 436 DGTSTAVPETELPLVPVT---KDIRPSGTGESPLANLTDWLEVTREDGVKGRRETNTMP 492
Qy 369 GWGLAKSY-----PADLASYPVDVLOTAN-----HL---QAWISFSY-- 404
Db 493 QWAGSSWYLYRYIDPHNTEKLADEDLKQWLPVDIVVGGAEHVLHLLYARFWHKLFDL 552
Qy 405 ---KGKPYMSVY 414
Db 553 GWVTEKPEFQKLF 565

RESULT 8

US-09-758-282B-155
; Sequence 155, Application US/09758282B
; Patent No. 6635463
; GENERAL INFORMATION:
; APPLICANT: Ma, Wu-Po
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamicheva, Natalie E.
; APPLICANT: Allawi, Hatim T.
; APPLICANT: Schaefer, James J.
; APPLICANT: Neri, Bruce P.
; TITLE OF INVENTION: Enzymes for the Detection of Nucleic Acid Sequences
; FILE REFERENCES: FORS 04931
; CURRENT APPLICATION NUMBER: US/09/758,282B
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 09/577,304
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 835
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-758-282B-155

Query Match

Best Local Similarity 3.9%; Score 90.5; DB 4; Length 835;

Matches 91; Conservative 41; Mismatches 130; Indels 99; Gaps 23;

Qy 86 TRFGLPELSFNYSKSLRF-----AFLPLGSLTGKDDPNT 123

Db 36 TSGEFPQVYGFARSLKALKEDGQAVVVDKAPFRHEAYKAGRAPTFDFFR 95

Qy 124 Q-AIRVLODLKAMV-----PGLDLEWFDHFTKALVVSE-----EARTLL-DRDIEIFVK 173

Db 96 OLALVKRLVDLLGLVLEAPGYEA---DVLGTLAKAREGVEVILGDRDF-----FO 148

QY 174 TQNKLAADLEPSGDIVLTKYIYIPRIKSIATGTPKRLM-FDAIKA--ADKFKVA-----225
Db 149 LLSEKVSLLPDG-----TLVTPKQVQEKYGVPPERWVDFRALTGDRSDNIPGVAGIGEK 203
QY 226 TPLAILEEFIAERAPTLGLHFLSCDLVXPSESRKIVYC-----MERQLDLASTEGIWTNG 281
Db 204 TALRLAELW--GSVENLLKNL-----DRVKPDSLRKIEAHLEDHLSDLRARTDLPLEV 258
QY 282 -----RRNDPETLDGDA-LRELWQLLPVTEGLCPNCFYE-----PGTSPQEQLPF---328
Db 259 DFKALRRTPD-LEGLRAFLEEL-----EFGSLLEHFGLLGGEKPREAPWPPP 306
QY 329 ---IINFTLSPKSALPEPOIYFPAFGQ-----NDKTIAGLATFFESRGWGLAKSPADL 381
Db 307 EGAFVGLLSRKEPMMWAEILLALAAASEGRVHRATSPVEALADLKEARGF--LAK-----DL 360
QY 382 A 382
Db 361 A 361
RESULT 9
US-09-758-282B-243
; Sequence 243, Application US/09758282B
; Patent No. 6635463
; GENERAL INFORMATION:
; APPLICANT: Ma, Wu-Po
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamicheva, Natalie E.
; APPLICANT: Allawi, Hatim T.
; APPLICANT: Schaefer, James J.
; APPLICANT: Neri, Bruce P.
; TITLE OF INVENTION: Enzymes for the Detection of Nucleic Acid Sequences
; FILE REFERENCE: FORS 04931
; CURRENT APPLICATION NUMBER: US/09/758,282B
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 09/577,304
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 243
; LENGTH: 835
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-758-282B-243
Query Match 3.9%; Score 90.5; DB 4; Length 835;
Best Local Similarity 25.2%; Pred. No. 3.2;
Matches 91; Conservative 41; Mismatches 130; Indels 99; Gaps 23;

QY 86 TRFGLPFLSPFNYSKLLRF-----AFPLGLSLTGKDDPFWT 123
Db 36 TSGEPVQVYGFARSLLKALKEDGOAVVVFDAKAPSPRHEAYEAYKAGRAPTDFPR 95
QY 124 Q-AIRPVLDLQKAMV-----PGLDLEWFDHFTKALVVSSE-----EARTLL-DRDIBIPVK 173
Db 96 QLALVXELVDLLGLVRLAEGYEA---DDVLGTAKKAEREGVEVRLTGDRDF----FQ 148
QY 174 TQNKLAADLEPSGDIVLTKYIYIPRIKSIATGTPKRLM-FDAIKA--ADKFKVA-----225
Db 149 LLSEKVSLLPDG-----TLVTPKQVQEKYGVPPERWVDFRALTGDRSDNIPGVAGIGEK 203
QY 226 TPLAILEEFIAERAPTLGLHFLSCDLVXPSESRKIVYC-----MERQLDLASTEGIWTNG 281
Db 204 TALRLAELW--GSVENLLKNL-----DRVKPDSLRKIEAHLEDHLSDLRARTDLPLEV 258
QY 282 -----RRNDPETLDGDA-LRELWQLLPVTEGLCPNCFYE-----PGTSPQEQLPF---328
Db 259 DFKALRRTPD-LEGLRAFLEEL-----EFGSLLEHFGLLGGEKPREAPWPPP 306

QY 329 ---IINFTLSPKSALPEPOIYFPAFGQ-----NDKTIAGLATFFESRGWGLAKSPADL 381
Db 307 EGAFVGLLSRKEPMMWAEILLALAAASEGRVHRATSPVEALADLKEARGF--LAK-----DL 360
QY 382 A 382
Db 361 A 361
RESULT 10
US-08-658-136-5
; Sequence 5, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMING, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-658-136-5
Query Match 3.9%; Score 90; DB 3; Length 4302;
Best Local Similarity 20.4%; Pred. No. 51;
Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;

QY 24 LSNKDHTK--WVYST-----APMFATMMAGAGYDVHAQYKFLCIHR-----EVIIPALG 70
Db 2487 LTTKVHECTGWDHDAEDAGAPLVYALLRRRCQGCE--FCVYKGSLSVGVLPFGFR 2544
QY 71 PYPKQGMHWSHLTRFGLPFELSFNYSKSLRFAPEPLGSLTGTKDDPFWTQAIRPV 130
Db 2545 PHFEVGLAVVQDQGAHVVAL-----NRSLATLPEPNGSATG-----LTWVL 2588
QY 131 QDLKAMV-PGI-----DLEWFDHFTKALVVSSEEBARTLLDRDIEIPVFKTNK-----177
Db 2599 HGLTASVLPGLLRADPOHVEISLALVTVNEVERALDVAAE-PKHQRHQAIRKNIT 2647
QY 178 -----LAADLEPSGDIVLTKYIYIPRIKSIATGTPKRLMFDKA 217
Db 2648 ETLVSLRVHTVDDIQIAAALAAQCGWPSRELVCRSCLKQTLHLK-----EAWML--ILQ 2699

FILING DATE: 23-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9507766.5
FILING DATE: 13-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326470.3
FILING DATE: 24-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Ph.D., Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/53313 (MRC-006xxx)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 345-9100
TELEFAX: (617) 345-9111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-422-582-8

Query Match 3.9%; Score 90; DB 4; Length 4302;
Best Local Similarity 20.4%; Pred. No. 51;
Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;

QY 24 LSNKDHK--WVYST---APMFATMAGAGYDVHAQYKFLCIHR-----EVIIPALG 70
DB 2487 LTTKVHPECTGWDHDAEDAGAPLVALLRRCRQHCCE--FCVYKGLSSYGAVLPQFR 2544
QY 71 PYPEKGQPMHWSHLTRFGLPFELSFNYSKSLRFAFEPGLSLGTGKDDPFNTQAIRPVL 130
DB 2545 PHFEVGLAVVQDQGAUVVAL-----NRS LAITLPEPNSGATG-----LTWVL 2588
QY 131 QDLKAMV-PGL-----DLEWDFHTKALVSEEEARTLLDRDIEIPVKTQNK-----177
DB 2589 HGLTASVLPGLLRQADPQHVIYSALVTVLNEYERALDVAAE--PKHERQHRAQIRKMIT 2647
QY 178 -----LAADLEPSGDIIVKTYIYPRIKSIATGTPKERLMFDAIKA 217
DB 2648 ETIVSLRVHTVDDIQIQAALAOCMGPSRELVCSCCLKQTLHLK-----EAMML--ILQ 2699
QY 218 ADKFGKVATPLAILEEFIAERAPTLG---HFLSCDLVKPSESRI-----KYC 263
DB 2700 AETTAGTVTPTAIGDSIL-----NITGDLIHLASSDVRAFPQPSSELGAESPSRMVASQAYN 2754
QY 264 M-----EROLDIASIEGIWTLNGRNDPDTLGDALRELWQLLPVTEGL 308
DB 2755 LTSALMRILMRSEVLNNEEPLTAGEIV--AQGKSDPRSL-----LCYGG 2798
QY 309 CPLPNCFY-----PGTSPOEQLPFI-----NFTLSPKSALPEQIYFP 348
DB 2799 APGPGCHFSIPEAFSGALANLSDVQLIFLVDSPNPPFGYISNYTVSTKVASMAFQTQAG 2858
QY 349 AFGQNDKTIAGELATF-----FESRGWGLAKSYPADL-----ASYYPDVDLQTN---394
DB 2859 AQPIERLASERAITVKVPNNSDWAARGHRSANSVVPQASVGAVVTLDSSNPA 2918
QY 395 --HLQAWIS-----FSYKGGKPYMSYLLHT 417
DB 2919 GLHLQNLNYLLDGHVLSERPEPVLYVLHS 2948

RESULT 13
US-09-052-262-8
Sequence 8, Application US/09052262
Patent No. 6656681
GENERAL INFORMATION:
APPLICANT: Harris et al.
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
NUMBER OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.
STREET: One Financial Center
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk, 3.50 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/052,262
FILING DATE: Concurrently herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/422,582
FILING DATE: 14-April-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9507766.5
FILING DATE: 13-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9411900.5
FILING DATE: 14-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB FCT/GB94/02822
FILING DATE: 23-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326470.3
FILING DATE: 24-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Ph.D., Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/74118
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 345-9100
TELEFAX: (617) 345-9111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-052-262-8

Query Match 3.9%; Score 90; DB 4; Length 4302;
Best Local Similarity 20.4%; Pred. No. 51;
Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;

QY 24 LSNKDHK--WVYST---APMFATMAGAGYDVHAQYKFLCIHR-----EVIIPALG 70
DB 2487 LTTKVHPECTGWDHDAEDAGAPLVALLRRCRQHCCE--FCVYKGLSSYGAVLPQFR 2544
QY 71 PYPEKGQPMHWSHLTRFGLPFELSFNYSKSLRFAFEPGLSLGTGKDDPFNTQAIRPVL 130
DB 2545 PHFEVGLAVVQDQGAUVVAL-----NRS LAITLPEPNSGATG-----LTWVL 2588
QY 131 QDLKAMV-PGL-----DLEWDFHTKALVSEEEARTLLDRDIEIPVKTQNK-----177
DB 2589 HGLTASVLPGLLRQADPQHVIYSALVTVLNEYERALDVAAE--PKHERQHRAQIRKMIT 2647
QY 178 -----LAADLEPSGDIIVKTYIYPRIKSIATGTPKERLMFDAIKA 217
DB 2648 ETIVSLRVHTVDDIQIQAALAOCMGPSRELVCSCCLKQTLHLK-----EAMML--ILQ 2699
QY 218 ADKFGKVATPLAILEEFIAERAPTLG---HFLSCDLVKPSESRI-----KYC 263
DB 2700 AETTAGTVTPTAIGDSIL-----NITGDLIHLASSDVRAFPQPSSELGAESPSRMVASQAYN 2754
QY 264 M-----EROLDIASIEGIWTLNGRNDPDTLGDALRELWQLLPVTEGL 308
DB 2755 LTSALMRILMRSEVLNNEEPLTAGEIV--AQGKSDPRSL-----LCYGG 2798
QY 309 CPLPNCFY-----PGTSPOEQLPFI-----NFTLSPKSALPEQIYFP 348

Db 2799 APGCGHFSIPEAFSGALANLSDVQLIFLVDNPPFFGVISNYTVSTKVASMAFQTQAG 2858
QY 349 AFQNDKXTIAEGLATP-----FESRGWGLAKSYPADL-----ASYYPDVLQITAN--- 394
Db 2859 AQPIERLASERAITVKVNNSDWAARGHRSSANSANSVVOPOASVGAVVTLDSNPAA 2918
QY 395 --HLQAWIS-----FSYKGGKPYMSVYLHT 417
Db 2919 GLHLQALNYTLDDGHYLSEPEPYLAVYLS 2948
RESULT 14
US-08-460-751-2
; Sequence 2, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Glucksmann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7638-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/8741
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4303 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-460-751-2
Query Match 3.9%; Score 90; DB 2; Length 4303;
Best Local Similarity 20.4%; Pred. No. 51;
Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;
QY 24 LSKNDHTK--WNYST----APMFATMAGAGYDVHQAQKFLCIHR-----EVIIPALG 70
Db 2487 LTTKVHFECTGHWDAEDACAPLVYALLRRRCQGHCEE--FCVYKGLSSYCAVLPPGFR 2544
QY 71 PYPERKQPMHWKSHLTRFGLPELGFNSKSLRFAFPLGLSLTGTKDDPNTQAIRPVL 130
Db 2545 PHEVGLAVVVOQDLGAAVVAL-----NRSIAITLPEPNSGATG-----LTVWL 2588
QY 131 QDLKAMV-PGL-----DLEWDFHTKALVYVSEBEARTLLDRDTEIFVFKTNK----- 177
Db 2589 HGTATSVLPGALROADPCHVIFYSIALVTWLVNFEVRAIDVAAF--EKHFEHRAOTIRKNT 2647

QY 178 -----LAADLEPSGDIVLKYIYPRIKSIATGTPKERMFDKA 217
Db 2648 ETVLSLVHTVDIQQIATAALAQCMPSELVCRSLKQTLHKL-----EAMML--ILQ 2699
QY 218 ADKFGKVATPLALEEFIAERAPTLG---HFISCDLVKPSERI-----KVYC 263
Db 2700 AETTAGTWTETAGDSIL-----NITGDLIHLASSDVRAPOPSSELGAESPSRMVASQAVN 2754
QY 264 M-----ERQLDLASIEGIWTNGRRNDPETHLGDLDALRELWQLLPVTEGL 308
Db 2755 LTSALMRILMRSRVLNDEEPLTAGEEV--AQGRSDPRSL-----LCYGG 2798
QY 309 CPLENCFYE-----PGTSPQEQLPFI-----NFTLSPKSALEPEPQIYFP 348
Db 2799 APGCGHFSIPEAFSGALANLSDVQLIFLVDNPPFFGVISNYTVSTKVASMAFQTQAG 2858
QY 349 AFQNDKXTIAEGLATP-----FESRGWGLAKSYPADL-----ASYYPDVLQITAN--- 394
Db 2859 AQPIERLASERAITVKVNNSDWAARGHRSSANSANSVVOPOASVGAVVTLDSNPAA 2918
QY 395 --HLQAWIS-----FSYKGGKPYMSVYLHT 417
Db 2919 GLHLQALNYTLDDGHYLSEPEPYLAVYLS 2948

RESULT 15
US-09-052-469-6
; Sequence 6, Application US/09052469
; Patent No. 6380360
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,469
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 06/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA: GB PCT/GB94/02822
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 6

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-052-469-6

Query Match          3.9%; Score 90; DB 4; Length 4339;
Best Local Similarity 20.4%; Pred. No. 51;
Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;

QY      24 LSNKDHK--WVYST-----APMFATMMAGAGYDVHQAQKFLCIHR-----EVIIPALG 70
Db      2444 LTTKVHECTGWHDAEDAGAPLYALLRRRCQGCHCE--FCVYKGSLSYGAIVPPGFR 2501
QY      71 PYPEKQPMHMKSHLTRFGLPELSFNYSKSLRFAFEPLGSLTGTGKDDPENTQAIRFVL 130
Db      2502 PHFEVGLAVVVDQLGAUVVAL-----NRSIAITLPEENGSGATG-----LTVWL 2545
QY      131 QDLKAMV-PGL-----DLEWFDHFTKALVSVSEBARTLDRDIEIPVKTKQNK----- 177
Db      2546 HGLTASVLPGLLRQADPOHVIYSYSLAVTVLNEYERALDVAAE-PKHERQHEAQIRKNIT 2604
QY      178 -----LAADLEPSGDIVLKTVIYPRIKSIATGTPKERLMFDAIKA 217
Db      2605 ETIVSLRVHTVDDIQIAAALAQCMGPFRELVCRSCLKQTLHKL-----EAMML--ILQ 2656
QY      218 ADKFGKVATPLALEEFIAERAPTLG---HFLSCDLVKPSESRI-----KYVC 263
Db      2657 AETTAGTVPTAIGDSIL-----NITGDLIHASSDVRAPOPSSELGAESPSPRMVASQAYN 2711
QY      264 M-----EROLDIASIEGIWTLNGRNDPETLDGLDALRELWQLLPVTEGL 308
Db      2712 LTSALMRIILMRSRVLNEEPLTLAGEEIV--AQKRSDDPRSL-----LCYGG 2755
QY      309 CPLPNCFYE-----PQTSFQEQLPFI-----NFTLSPKSALEPEQIYPP 348
Db      2756 APGGCHFSIPEAFSCALANLSDVQLIFLVDNPPPGYISNYTVSTKVASMAFQTQAG 2815
QY      349 AFCQNDKTIAGLATP-----FESRGWGGLAKSYPADL-----ASYYPDVDLQTAN--- 394
Db      2816 AQIPIERLASERAITVKVPNNSDWAARGHRSSANSANSVVQFQASVGAVVTLDSNPAA 2875
QY      395 --HLQAWIS-----FSYKGGKKPYMSVYLHT 417
Db      2876 GLHLQLNYTLTDGHYLSSEPEPYLAVYLHS 2905
```

Search completed: July 22, 2004, 18:46:58
Job time : 21 secs

Sequence 19, Appl
Sequence 3347, Ap
Sequence 168, App
Sequence 72824, A
Sequence 66046, A
Sequence 43, Appl
Sequence 36, Appl
Sequence 121262,
Sequence 3386, Ap
Sequence 149090,
Sequence 54706, A
Sequence 701, App
Sequence 23, Appl
Sequence 30, Appl
Sequence 10341, A
Sequence 241638,
Sequence 38939, A
Sequence 159892,
Sequence 3, Appli
Sequence 74920, A
Sequence 11, Appl
Sequence 20151, A
Sequence 7965, Ap
Sequence 7966, Ap
Sequence 308, App
Sequence 744, App
Sequence 13226, A
Sequence 795, App
Sequence 65022, A

16 102 4.4 1016 9 US-09-986-224-19
17 101.5 4.4 930 15 US-10-369-493-3347
18 29.5 4.3 3234 15 US-10-093-463-168
19 98.5 4.2 507 12 US-10-425-114-72824
20 98.5 4.2 769 12 US-10-282-122A-66046
21 98 4.2 1939 12 US-10-152-886-43
22 97.5 4.2 344 12 US-10-617-038-36
23 96.5 4.2 1095 16 US-10-437-963-121262
24 95.5 4.1 503 15 US-10-369-493-3386
25 95 4.1 876 16 US-10-437-963-149090
26 94.5 4.1 430 12 US-10-282-122A-54706
27 94.5 4.1 430 13 US-10-289-762-701
28 94.5 4.1 453 16 US-10-448-871A-23
29 94.5 4.1 578 8 US-08-808-031A-30
30 94 4.0 466 14 US-10-156-761-10341
31 93.5 4.0 818 12 US-10-424-599-241638
32 93.5 4.0 818 12 US-10-425-114-38939
33 93 4.0 714 16 US-10-437-963-159892
34 93 4.0 2492 9 US-09-991-358-3
35 92.5 4.0 585 12 US-10-282-122A-74920
36 92.5 4.0 862 14 US-10-267-989-11
37 92.5 4.0 3250 15 US-10-369-493-20151
38 92 4.0 443 12 US-10-335-977-7965
39 92 4.0 448 12 US-10-335-977-7966
40 92 4.0 491 10 US-09-934-455-308
41 92 4.0 491 12 US-10-225-066A-744
42 92 4.0 491 15 US-10-374-780A-2474
43 92 4.0 597 9 US-09-815-242-13226
44 92 4.0 608 12 US-10-087-192-795
45 91.5 3.9 570 12 US-10-282-122A-65022

ALIGNMENTS

RESULT 1
US-10-099-704-2
; Sequence 2, Application US/10099704
; Publication No. US20020197682A1
; GENERAL INFORMATION:
; APPLICANT: Christensen, Bjorn Eggert
; APPLICANT: Mollgaard, Henrik
; APPLICANT: Kaasgaard, Svend
; APPLICANT: Lehmebeck, Jan
; TITLE OF INVENTION: Methods for producing polypeptides in
; FILE OF INVENTION: aspergillus mutant cells
; FILE REFERENCE: 4483.200-US
; CURRENT APPLICATION NUMBER: US/10/099,704
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/472,364
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/139,593
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: 60/117,396
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: PA 1999 00745
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: PA 1998 01726
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer 5956
US-10-099-704-2

Query Match 100.0% Score 2322; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.5e-224;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2322	100.0	437	13	US-10-099-704-2
2	402	17.3	330	15	US-10-369-493-12970, A
3	346.5	14.9	430	15	US-10-369-493-13332
4	313.5	13.5	249	15	US-10-369-493-12523
5	293.5	12.6	379	15	US-10-369-493-3945
6	114	4.9	577	16	US-10-437-963-182042
7	110	4.7	477	12	US-10-425-114-69165
8	104.5	4.5	393	16	US-10-437-963-122089
9	104	4.5	524	15	US-10-369-493-2592
10	103.5	4.5	880	14	US-10-059-909-12
11	102	4.4	521	9	US-09-848-035-6
12	102	4.4	521	9	US-09-986-224-6
13	102	4.4	605	15	US-10-407-866-74
14	102	4.4	635	15	US-10-161-927-4
15	102	4.4	662	15	US-10-407-866-101

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:46:15 ; Search time 45 Seconds
(without alignments)
3041.075 Million cell updates/sec

Title: US-10-099-704-2

Perfect score: 2322

Sequence: 1 MSISKAATLLPKPFVLSQ.....FEAFSAQAQVAMCHDGHNP 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1288442 segs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

QY 1 MEISKKAATLLPKPFVYLSQALNSKDHKWKYSTAPMFATMMAGAGYDVHQAQYKFLCI 60
DB 1 MEISKKAATLLPKPFVYLSQALNSKDHKWKYSTAPMFATMMAGAGYDVHQAQYKFLCI 60
QY 61 HREVIIPALGPYKPGPMKWSHLTRFGLPFELSFNYSKSLRFAFEPGLSGTGTGKDDP 120
DB 61 HREVIIPALGPYKPGPMKWSHLTRFGLPFELSFNYSKSLRFAFEPGLSGTGTGKDDP 120
QY 121 FNTQAIRPVLDLKAAMVPGDLLEWDFHFTKALVVSSEEARLTDRIEIPVKTQNKLA 180
DB 121 FNTQAIRPVLDLKAAMVPGDLLEWDFHFTKALVVSSEEARLTDRIEIPVKTQNKLA 180
QY 181 DLPSGDIVLKTYYIPRIKSIATGTPKERLMFDAIKAADKFGKVATPLAILBEEFAERAP 240
DB 181 DLPSGDIVLKTYYIPRIKSIATGTPKERLMFDAIKAADKFGKVATPLAILBEEFAERAP 240
QY 241 TLGHFLSCDLVKPESRKYCMERQDLASIEGWTINCRNDPDTLGDALRELWQ 300
DB 241 TLGHFLSCDLVKPESRKYCMERQDLASIEGWTINCRNDPDTLGDALRELWQ 300
QY 301 LLPVTEGLCLPNCVPEGTSPQEQLPFIINFTLSPKSLPPEQIYFPAGQNDKTIABG 360
DB 301 LLPVTEGLCLPNCVPEGTSPQEQLPFIINFTLSPKSLPPEQIYFPAGQNDKTIABG 360
QY 361 LATPFESRGWGLAKYPADLASYPDVLQTNHLQAWISFSYKGGKPYNSVYLHTEFA 420
DB 361 LATPFESRGWGLAKYPADLASYPDVLQTNHLQAWISFSYKGGKPYNSVYLHTEFA 420
QY 421 FSAACAEVAMCHDGNP 437
DB 421 FSAACAEVAMCHDGNP 437
RESULT 2
US-10-369-493-12970
; Sequence 12970, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12970
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; NAME/KEY: unsure
; LOCATION: (1)...(430)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12970
Query Match 17.3%; Score 402; DB 15; Length 330;
Best Local Similarity 31.3%; Pred. No. 2e-31;
Matches 105; Conservative 61; Mismatches 125; Indels 44; Gaps 12;
QY 10 LKPK-----PFYVLSQALNSKDHKWKYSTAPMFATMMAGAGYDVHQAQYKFLCIHREY 64
DB 1 LKPKPISTQPTLLTSLYLPFTNEAQWLDGSLFSLRQASQYTTGQYKHLFFPHY 60
QY 65 IIPALGPYKPGPMKWSHLTRFGLPFELSFNYSKSLRFAFEPGLSGTGTGKDDP--D 119
DB 61 LKPLGLTYPT-----WFSIVSRQGLAIEYSLNFOASGNPVSRAFEPLSVASG-KDAAD 114
QY 120 PNTQAIRPVLDLKAAMVPGDLLEWDFHFTKALVVSSEEARLTDRIEIPVKTQNK 178
DB 115 PLNKRTEDLLAGLVQGLEGDMTLFDHVRTITFISDEETALENPDAY--GKHTIAL 172

QY 179 AADLEPSGD-IVLAKTYIYPRIKSIATGTPKERLMFDAI-KAADKFGKVATPLAILBEEFA 236
DB 173 GFDWK--GDNVTVKCYNHPKWSLATGVPVAKLIRDSLERIKQOF-DCRQALBLVDIYME 229
QY 237 ERAPTLGHFLSCDLVKPESRKYCMERQDLASIEGWTINCRNDPDTLGDALR 296
DB 230 DSGSWDLRTFIAMDVCLAQTELKIYGIINEVSLGKVEELWTMGRLNDETTLEGLSLR 289
QY 297 ELWOLL-----PVTGLCLPL 311
DB 290 RLLHLLVKNDDRFLSKGBEKLEYGPTVDGFLPL 324
RESULT 3
US-10-369-493-13332
; Sequence 13332, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13332
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; NAME/KEY: unsure
; LOCATION: (1)...(430)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-13332
Query Match 14.9%; Score 346.5; DB 15; Length 430;
Best Local Similarity 26.3%; Pred. No. 1.1e-25;
Matches 120; Conservative 55; Mismatches 177; Indels 105; Gaps 15;
QY 17 VLSQALNSKDHKWKYSTAPMFATMMAGAGYDVHQAQYKFLCIH----- 61
DB 7 VLSQTLPSRGPDVDAMWQLTGRHLAVLLDAAAYPIEKQYECCLLYHYHVAVCLTTSYKVT 66
QY 62 ---REVLIIPALGPYKPGPMKWSHLTRFGLPFELSFNYS-----KSLRFAFEPGLSG 113
DB 67 DXLEVXAPYLGPAPEGASPTTKWSMLQDGTFFFSKWNPNPGGPDVRFGLSEIGPM 126
QY 114 TGTKDDPNTQAIRPVLDLKAAMVPGDLLEWDFHFTKALVVSSEEARLTDRIEIPVK 173
DB 127 AGTSLDPLNHLAMKEILYKLSAVPGSGLTWTHF-----LATLFDH-----YA 171
QY 174 TQNKLAADLEPS-----GDVILKTYIYR-----IKSIATG 204
DB 172 KYTQKAATMGSSITGTSLSWLSLEFQRKSTGLKTYFHPKLDQQQAFLDIPSWAESFRGLHPN 231
QY 205 TPKERLMFDAIKAADKFGKVATPLAILBEEFAERAPTLGHFLSCDLVKPESRKYCM 264
DB 232 SPRTAVHEFL-STNPEGKLLKPPC-----LSVNCSPAKARIKWYN 273
QY 265 BRQDLASIEGWTINCRNDPDTLGDALRELWOLL-PVTBGLCLPNCVPEGTGTSQ 323
DB 274 SPHTNFRAIREINTLGGRIADTET--RTKQFSELNLLKTVTRGARXFFPRDF-RVSIQ 330
QY 324 EQ-----LPIFIN-----FTLSPKSLPPEQIYFPAGF--ONDKTIAEGLATFFES 367
DB 331 ORQHNPQLDADPMLKGCYVFFDIAPGNLPAIKYFFPVNRHCRNDLAVTQNLNEWLES 390

```

/ CLONING PROJECT 1. US/10/437, 963
/ CURRENT APPLICATION NUMBER: US/10/437, 963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 182042
/ LENGTH: 577
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURES:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_79266C.1.pep
US-10-437-963-182042

Query Match          4.9%; Score 114; DB 16; Length 577;
Best Local Similarity 24.8%; Pred. No. 0.045;
Matches 110; Conservative 53; Mismatches 155; Indels 126; Gaps 29;

Qy      62  REVIIIPALGPY--PEKGQPM-----HMKSHL-----TFGGLPFELSPNYSK 100
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
db      165  REWKRYAINPVANPEPCGNATOFDPTTANVADGHFWMLVGGLKGSLGLAYLYRSRDFK 224

```


RESULT 10
US-10-059-909-12
/ Sequence 12, Application US/10059909
/ Publication NO. US20030074693A1
/ GENERAL INFORMATION:
/ APPLICANT: Cahoon, Edgar B.
/ APPLICANT: Kinney, Anthony
/ APPLICANT: Klein, Theodore
/ APPLICANT: Lee, Jian Ming
/ APPLICANT: Pearlstein, Richard
/ APPLICANT: Rafalski, J. Antoni
/ APPLICANT: Shen, Jennie
/ APPLICANT: Thorpe, Cathy
/ APPLICANT: Tingey, Scott
/ APPLICANT: Weng, Zude
/ TITLE OF INVENTION: Plant Lipoxigenases
/ FILE REFERENCE: B81333 US CIP
/ CURRENT APPLICATION NUMBER: US/10/059,909
/ CURRENT FILING DATE: 2002-01-29
/ PRIOR APPLICATION NUMBER: 09/501,422
/ PRIOR FILING DATE: 2000-02-09
/ PRIOR APPLICATION NUMBER: 60/119,597
/ PRIOR FILING DATE: 1999-02-10

```

; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/201,464
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 521
; TYPE: PAT
; ORGANISM: Homo sapiens
US-09-848-035-6

Query Match
Best Local Similarity    4.4%; Score 102; DB 9; Length 521;
Matches      81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;

QY          79 MWKSHL---TRFGPEELSNFYNSKLLEFAFEPGLSGTGTGK-DDPFNQAIRPVLODLK 134
           :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          67 LHWANGVLFQQRPFSYVFVLSCHKRYMKETTFAELISLDWPDFDAPIEEFMSQP--EKLL 124
           :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY          135 AVMPGLDLEWFPHFTKALVVSSEEAARTLLDRD-----IEIPVK-TQNKLAADLEPSGD 187
           :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          125 FIIDGFE-----EIIISERSGESLDDGSFCTDWYOELPVTVKILHSLKKELVPLAT 175
           :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

QY 188 --IVLKYIYPRKSI-----ATGTPKERLMFDAIKAAKFGKVATPLAILEE----- 233
Db 176 LLITIKTFVRDLKASLVNFCFVQITGTTGDDLRVFMRFHFDSDSSEVEKILQQLRKNETL 235
QY 234 FIABRPTLLGHFLSC-----DLVKPSESRIKYVC-----MERQLDLA----- 271
Db 236 FHSCSAPMVCWTVCCLQPKVRYDLSQITQTTSLYAYFFSNLSTAEVDLADDSWPG 295
QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 296 QWRALCSLAIEGLWSMNFNFKEDTEIEGLEVPFIDSLYEFNLIQKINDCGGCTTFTHLS 355
QY 312 -----PNCFYEPGTSPOQLPFIINFITLSPKLSALPEPOLYFPAPGQNDKTI 357
Db 356 FOEFAAMS FVLEBEPFPFHSTKPOE-MQMLLQHVLLDKAEYWTVPVLP-FFGLLNKNI 413
QY 358 AEGL 361
Db 414 AREL 417

RESULT 12
US-09-986-224-6
; Sequence 6, Application US/09986224
; Patent No. US20020150920A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; APPLICANT: Wang, Weiye
; TITLE OF INVENTION: NOVEL MOLECULES OF THE NBS/LRR PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-333001
; CURRENT APPLICATION NUMBER: US/09/986,224
; 2001-10-22
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/848,035
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 60/201,464
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-224-6

Query Match 4.4%; Score 102; DB 9; Length 521;
Best Local Similarity 22.3%; Pred. No. 0.63;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
QY 79 MHWKSHL---TRFGLPELSEFNYSKSLRFAFPGLSLGTGK-DDPFTQAIRPVLDLQK 134
Db 67 LHWANGVLFOQRFYSYVFLSCHIRYMKETTFAEILSLDWPDPDAPIEFMSQP--EKLL 124
QY 135 AMVPGDLEWFDHFTKALVVSBEARTLDRD-----TEIPVK-TQNKLAADLSPGD 187
Db 125 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTIKILHSLKKEVLPLAT 175
QY 188 --IVLKYIYPRKSI-----ATGTPKERLMFDAIKAAKFGKVATPLAILEE----- 233
Db 176 LLITIKTFVRDLKASLVNFCFVQITGTTGDDLRVFMRFHFDSDSSEVEKILQQLRKNETL 235
QY 234 FIABRPTLLGHFLSC-----DLVKPSESRIKYVC-----MERQLDLA----- 271
Db 236 FHSCSAPMVCWTVCCLQPKVRYDLSQITQTTSLYAYFFSNLSTAEVDLADDSWPG 295
QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 296 QWRALCSLAIEGLWSMNFNFKEDTEIEGLEVPFIDSLYEFNLIQKINDCGGCTTFTHLS 355
QY 312 -----PNCFYEPGTSPOQLPFIINFITLSPKLSALPEPOLYFPAPGQNDKTI 357

QY 358 AEGL 361
Db 414 AREL 417

RESULT 13
US-10-407-866-74
; Sequence 74, Application US/10407866
; Publication No. US20040002593A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides, and Methods of Use
; FILE REFERENCE: 66654-10 (LJ 5755)
; CURRENT APPLICATION NUMBER: US/10/407,866
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: US 60/370,538
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-407-866-74

Query Match 4.4%; Score 102; DB 15; Length 605;
Best Local Similarity 22.3%; Pred. No. 0.79;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
QY 79 MHWKSHL---TRFGLPELSEFNYSKSLRFAFPGLSLGTGK-DDPFTQAIRPVLDLQK 134
Db 232 LHWANGVLFOQRFYSYVFLSCHIRYMKETTFAEILSLDWPDPDAPIEFMSQP--EKLL 289
QY 135 AMVPGDLEWFDHFTKALVVSBEARTLDRD-----TEIPVK-TQNKLAADLSPGD 187
Db 290 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTIKILHSLKKEVLPLAT 340
QY 188 --IVLKYIYPRKSI-----ATGTPKERLMFDAIKAAKFGKVATPLAILEE----- 233
Db 341 LLITIKTFVRDLKASLVNFCFVQITGTTGDDLRVFMRFHFDSDSSEVEKILQQLRKNETL 400
QY 234 FIABRPTLLGHFLSC-----DLVKPSESRIKYVC-----MERQLDLA----- 271
Db 401 FHSCSAPMVCWTVCCLQPKVRYDLSQITQTTSLYAYFFSNLSTAEVDLADDSWPG 460
QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 461 QWRALCSLAIEGLWSMNFNFKEDTEIEGLEVPFIDSLYEFNLIQKINDCGGCTTFTHLS 520
QY 312 -----PNCFYEPGTSPOQLPFIINFITLSPKLSALPEPOLYFPAPGQNDKTI 357
Db 521 FOEFAAMS FVLEBEPFPFHSTKPOE-MQMLLQHVLLDKAEYWTVPVLP-FFGLLNKNI 578
QY 358 AEGL 361
Db 579 AREL 582

RESULT 14
US-10-161-927-4
; Sequence 4, Application US/10161927
; Publication No. US20030235821A1
; GENERAL INFORMATION:
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Miller, Charles E.
; APPLICANT: Hjalt, Tord

APPLICANT: Baumgartner, Jason C.
APPLICANT: Guo, Xiaojia
APPLICANT: Gangolli, Esha A.
APPLICANT: Vernet, Corine
APPLICANT: Padigar, Muralidhara
APPLICANT: Li, Li
APPLICANT: Pena, Carol E.A.
APPLICANT: Gorman, Linda
APPLICANT: Anderson, David W.
APPLICANT: Edinger, Shlomit R.
APPLICANT: Patturajan, Meera
APPLICANT: Stone, David J.
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
FILE REFERENCE: 21402-377 D (Cura 677 Other)
CURRENT APPLICATION NUMBER: US/10/161,927
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: 60/295,661
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/295,607
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/296,404
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: 60/296,418
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: 60/296,575
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: 60/297,414
PRIOR FILING DATE: 2001-06-11
PRIOR APPLICATION NUMBER: 60/297,567
PRIOR FILING DATE: 2001-06-12
PRIOR APPLICATION NUMBER: 60/298,528
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/325,685
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/299,133
PRIOR FILING DATE: 2001-06-18
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 190
SEQ ID NO 4
LENGTH: 635
TYPE: PRT
ORGANISM: Homo sapiens
US-10-161-927-4

Query Match 4.4%; Score 102; DB 15; Length 635;
Best Local Similarity 22.3%; Pred. No. 0.85;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
Qy 79 MHKSHL---TRGGLPELSFNYSKSLRFAPEPLGSLGTGK-DDPFTQAIRPVLODLK 134
Db 67 LHWANGVLFQORFSYFVLSCHKIRYMKETTFAEELISLDWPDFDAPIEEFMSQP--EKLL 124
Qy 135 AMVPGLDLEWDFHFTKALVSEBEARTLLDRD-----IEIPVK-TQNKLAADLEPSGD 187
Db 125 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTKILHLLKKELVPLAT 175
Qy 188 --IVLKYIYPRIKSI-----ATGTPKERMFDKAAKDFKGVATPLAILEE---- 233
Db 176 LLITIKTWFRDLKASLVNFCVQITGTDGLVYFMRHFDSDSEVEKILQOLRKNETL 235
Qy 234 FIAERAPTLIGHFLSC-----DLVKPSESRIKYC-----MERQLDLA----- 271
Db 236 FHSCAPMVCWVCSCLKQPKVRYDLSQITQTTSYAYFFSNLSTAEVDLADDSWPG 295
Qy 272 -----SIEGIWTLNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 296 QWRALCSLAIEGLWSMNFTEKDETEIEGLEVPFIDSLYEFNLIQKINDCGGCTTFTHLS 355
Qy 312 -----PNCFYEPGTSPOQLPFIINFTLSPKSALPEPQIYFPAGQNDXTI 357
Db 356 FQEFFAAMSFVLEBPPEPFPSTKPOB-MKMLQHVLLDKKEYWTPVVLFF-FFGLLNKNI 413

Qy 358 AEGL 361
Db 414 AREL 417
RESULT 15
US-10-407-866-101
; Sequence 101, Application US/10407866
; Publication No. US20040002593A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
; FILE REFERENCE: 66654-10(LJ 5755)
; CURRENT APPLICATION NUMBER: US/10/407,866
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: US 60/370,538
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-407-866-101
Query Match 4.4%; Score 102; DB 15; Length 662;
Best Local Similarity 22.3%; Pred. No. 0.9;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
Qy 79 MHKSHL---TRGGLPELSFNYSKSLRFAPEPLGSLGTGK-DDPFTQAIRPVLODLK 134
Db 232 LHWANGVLFQORFSYFVLSCHKIRYMKETTFAEELISLDWPDFDAPIEEFMSQP--EKLL 289
Qy 135 AMVPGLDLEWDFHFTKALVSEBEARTLLDRD-----IEIPVK-TQNKLAADLEPSGD 187
Db 290 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTKILHLLKKELVPLAT 340
Qy 188 --IVLKYIYPRIKSI-----ATGTPKERMFDKAAKDFKGVATPLAILEE---- 233
Db 341 LLITIKTWFRDLKASLVNFCVQITGTDGLVYFMRHFDSDSEVEKILQOLRKNETL 400
Qy 234 FIAERAPTLIGHFLSC-----DLVKPSESRIKYC-----MERQLDLA----- 271
Db 401 FHSCAPMVCWVCSCLKQPKVRYDLSQITQTTSYAYFFSNLSTAEVDLADDSWPG 460
Qy 272 -----SIEGIWTLNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 461 QWRALCSLAIEGLWSMNFTEKDETEIEGLEVPFIDSLYEFNLIQKINDCGGCTTFTHLS 520
Qy 312 -----PNCFYEPGTSPOQLPFIINFTLSPKSALPEPQIYFPAGQNDXTI 357
Db 521 FQEFFAAMSFVLEBPPEPFPSTKPOB-MKMLQHVLLDKKEYWTPVVLFF-FFGLLNKNI 578
Qy 358 AEGL 361
Db 579 AREL 582
Search completed: July 22, 2004, 18:51:49
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:41:47 ; Search time 17 Seconds
(without alignments)
2472.688 Million cell updates/sec

Title: US-10-099-704-2

Perfect score: 2322
Sequence: 1 MEISKKAATLLPKPFYVLSQ.....FEAFSAAGAEVAMCHDGNP 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	895.5	38.6	455	1 JC4338	tryptophan dimethyl
2	111.5	4.8	378	2 C64091	adenine glycosylas
3	111.5	4.8	555	1 SYHQWA	malate synthase [E
4	107.5	4.6	490	2 T23791	hypothetical prote
5	101	4.3	878	2 S74207	lipoxigenase [EC 1
6	101	4.3	1621	2 A82255	hypothetical prote
7	100	4.3	323	2 G87379	heptosyltransferas
8	99.5	4.3	486	2 JC4240	lanosterol 14alpha
9	99.5	4.3	503	2 JC4758	transcription init
10	98.5	4.2	503	2 G85432	GTP pyrophosphokin
11	98.5	4.2	769	2 D81828	hypothetical prote
12	97.5	4.2	344	2 F70922	aldehyde-ferradoxi
13	96.5	4.2	621	2 A71221	hypothetical prote
14	95.5	4.1	1073	2 S69079	genome polyprotein
15	95.5	4.1	1882	1 GNVVTR	histidyl tRNA synt
16	95	4.1	430	2 D86573	oligopeptide ABC t
17	95	4.1	486	2 B83909	histidine-tRNA lig
18	94.5	4.1	430	2 H72052	pol polyprotein -
19	94.5	4.1	896	1 GNLJQH	hypothetical prote
20	94	4.0	910	2 S76034	hypothetical prote
21	93.5	4.0	693	2 T33251	leucyl-tRNA synth
22	93	4.0	833	2 H95029	nonstructural poly
23	93	4.0	2492	1 MNWVTD	dihydroterocate sy
24	92.5	4.0	476	2 E69426	DNA-packaging prot
25	92.5	4.0	640	2 T13088	aldehyde dehydroge
26	92.5	4.0	862	1 A49346	F54F2.8 protein -
27	92	4.0	282	2 S44825	hypothetical prote
28	92	4.0	374	2 AD2802	periplasmic bindin
29	92	4.0	379	2 E97581	

30 92 4.0 443 2 F71929 heat shock protein
31 92 4.0 534 2 T04663 hypothetical prote
32 92 4.0 1017 2 S62435 probable glycine d
33 92 4.0 2492 1 C44213 nonstructural poly
34 91.5 3.9 466 2 T40150 probable glycylpep
35 91.5 3.9 585 2 AG0955 conserved hypothet
36 91.5 3.9 986 2 T41809 DNA polymerase orf
37 91.5 3.9 987 2 AF2296 hypothetical prote
38 91 3.9 399 2 S75523 mannoseyltransferas
39 91 3.9 470 2 S50083 photolyase - short
40 91 3.9 775 2 S28284 hypothetical prote
41 91 3.9 833 2 C97901 leucine-tRNA ligase
42 90.5 3.9 521 2 D64434 serine-tRNA ligase
43 90.5 3.9 586 2 JC2407 homothallic switch
44 852 3.9 S41886 DNA repair protein
45 90.5 3.9 1736 2 T05174 hypothetical prote

ALIGNMENTS

RESULT 1

JC4338
tryptophan dimethylallyltransferase (EC 2.5.1.34) - ergot fungus
N:Alternate names: dimethylallyl diphosphate:l-tryptophan dimethylallyltransferase; dim
C:Species: Claviceps purpurea (ergot fungus)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: JC4338
R:Tsal, H.F.; Wang, H.; Gebler, J.C.; Poulter, C.D.; Schardl, C.L.
Biochem. Biophys. Res. Commun. 216, 119-125, 1995
A:Title: The Claviceps purpurea gene encoding dimethylallyltryptophan synthase, the comm
A:Reference number: JC4338; MUID:96067540; PMID:7488077
A:Accession: JC4338
A:Molecule type: mRNA
A:Residues: 1-455 <TSA>
A:Cross-references: GB:L39640; NID:g1005417; PID:g1005418
A:Experimental source: ATCC 26245
C:Genetics:
A:Gene: dmaW
A:Introns: 396/2; 437/1
C:Function:
A:Description: catalyzes the synthesis of 4'-(gamma,gamma-dimethylallyl)-tryptophan fro
A:Pathway: ergot alkaloid biosynthesis
C:Superfamily: ergot fungus tryptophan dimethylallyltransferase
C:Keywords: alkaloid biosynthesis; transferase
F:113-117/Region: prenyl diphosphate binding #status predicted

Query Match 38.6%; Score 895.5; DB 1; Length 455;
Best Local Similarity 42.1%; Pred. No. 2.2e-64;
Matches 187; Conservative 75; Mismatches 151; Indels 31; Gaps 9;

QY 3 ISKKAATLLPKPFYVLSQALNSKDKTKWYVSTAPMEATWAGAGYDVHAQYKFLCIHR 62
DB 2 MTKAPATAV---YDTLSLLDFPNQEQRLMWHSTAPMFAAMLDTAGHNVHDQYRELGIK 58
QY 63 EVIIPALGPYEPKQPMH-WKSHLTRFGLPELSFNYSKSLRFAFEPLGLSTGTKDPPF 121
DB 59 KHIIFLGVYPAQK--HTWPSVLTRYGIPELSLNCILDSVVRYTFEPTTEHTGTGDDSY 116
QY 122 NTOATRPVLQDKAMVPGLDLEWPHFTKALVWSEEARLTILDRD-IEIPVKTONKLA 180
DB 117 NAFALTECIQKLVRIQPGIDMEWFSYFRNELVNLNATESARLGRNDSVNVQPIRTONKIAL 176
QY 181 DLEPSGD-IVLKYTYIPRIKSTATGTPKRLMFDAT-KAADKFGKVATPLAILEEFIAER 238
DB 177 DUK--GDRFALKVLYLPHLKSATGVSSHDLIFNSVRKLSQKHTSIQSFNVLCYVASR 234
QY 239 -----APTLLGHFLSCDLVVKSESRKIVYCHMERQDLIASIEGITWLNGR 283
DB 235 NPDPSNAAEAGVPSALRALRLISCGLVDPSKRIKTYLLEQTVSLTAMEDLWLGRR 294
QY 284 NDEPITLGDALRELWQLLPVTEGLCPLENCPEYEGTSPQEQLPFIINFTLSPKSALPEP 343

Db 295 TDSSTLNGLDMMRLHLLQIPSGFMKYPKPSDLKLGVEVPDEQLPSMWHYALHPDQMPER 354
Qy 344 QIYFPAFGNDKTTAEGLATFFSRGWLAKSPADLASYPDVLQANTHLQAWISFS 403
Db 355 QVYFTVFGMSDAGITNALATFFSRHGWYEMAKYRVFLEGSFPHNDPESLNYLHTVVSFS 414
Qy 404 YGKKKPYMSVYLHTE-----APS 422
Db 415 YRKNKPYLSVYLHSFETGQWPAFS 438

RESULT 2
C64091
adenine glycosylase - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 17-Nov-2000
C:Accession: C64091
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kexlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: C64091
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-378 <TIGR>
A:Cross-references: GB:U32760; GB:I42023; NID:gl573764; PIDN:AAC2418.1; PID:gl573768; T
C:Function:
A:Description: catalyzes the excision of adenine from a guanine-adenine mispair
C:Superfamily: A/G-specific adenine glycosylase
C:Keywords: DNA repair

Query Match 4.8%; Score 111.5; DB 2; Length 378;
Best Local Similarity 22.8%; Pred. No. 0.35;
Matches 79; Conservative 48; Mismatches 127; Indels 93; Gaps 21;

Qy 142 LEWFDHFTK-----ALVYSEBEARTLLDRDIEIPVF-----KTQKLA 180
Db 16 LAWYDFGRCHLEWQNKTLGYVWLVSEVMLQQQTQVATV-----IPYFERFKTFPNITA 69
Qy 181 DLSPSGDIVLKY-----IYPRKSIATGTPKER-----LMFDAKAAKFGKVATPL 228
Db 70 LANASQDEVLLHWTGLGYARAEHLKAAQKVRDEFNGFNFTNFQWALSGVGR-STAG 128
Qy 229 AILEEFIAERAPTLHGLFLSCDLVVKPSER---IKVCMEROLDIASIEGIWTLNLRND 285
Db 129 AILSSVLNQYPILDGN-----VKAVLARYFAVEGWSGEKKVE-----NRLWALT-EQVT 177
Qy 286 PET--LDGLDALRELHQLLPVTE-----GLCPUP-NCFYEPGTSPOBQFPFIINFTLSPKS 338
Db 178 PTTRVADFQNMMDIGAMVCMRTKPKRCDLPLMIDCLAYKNTN-WEKFP-----AKGPKK 231
Qy 339 ALPEPOLYPPAFQNDKTTAEGLATFFESRG-WGGLAKSYP--ADLASYPDVIDQTANH 395
Db 232 AMPEKTYTFLILSKNGKVCLEQR-----ENSLWGLGF-CFPOFEKXSLHLHFLAQEKVTH 286
Qy 396 LQWISFSYKGGKPYMSVYLHTEAFSA-----AAQEVAMCHDGHN 436
Db 287 YQWSPSFR-----HTFSHEHLDIHPYAEEMESTLQVEQAN 321

RESULT 3
SYQMA
malate synthase (EC 4.1.3.2), microbody - yeast (Pichia angusta)
C:Species: Pichia angusta
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
C:Accession: S09294
R:Bruinenberg, P.G.; Blaauw, M.; Kazemier, B.; AB, G. Yeast 6, 245-254, 1990
A:Title: Cloning and sequencing of the malate synthase gene from Hansenula polymorpha.
A:Reference number: S09294; MUID:90273778; PMID:2349836

A:Accession: S09294
A:Molecule type: DNA
A:Residues: 1-555 <BRU>
C:Comment: This enzyme, together with isocitrate lyase, is very important to the glyoxylate shunt.
C:Genetics:
A:Gene: MAS
C:Superfamily: malate synthase
C:Keywords: carbon-carbon lyase; coenzyme A; glyoxylate bypass; oxo-acid-lyase; tricarbo-

Query Match 4.8%; Score 111.5; DB 1; Length 555;
Best Local Similarity 23.5%; Pred. No. 0.61;
Matches 73; Conservative 39; Mismatches 89; Indels 109; Gaps 20;

Qy 16 YVLSQALNLSNKDHT-----KWWYSTAPMPATMAGAGYDVHAQYKFLCIHREVIIPAL 69
Db 269 YOLDEVLYALURDHSAGLNCGRWDY---MFSTI-----KELRNQKHILP-- 309
Qy 70 GPYPEKGQPMHWKSHLTRFRGLPFELSFNYSKSLRFAFE---PLGSLTGT---KDDP-F 121
Db 310 -----DRHQVTMTVPF--MTNYVKQLIKICHKGVHAMGGMAATIPKDDPEK 355
Qy 122 NQAIKRVLPQD-LKAMVPGDLLEWFDH---FTKALVYSEBEART-----LLORDIEIPVFK 173
Db 356 NAAAMEAVRQDKLEVLAGHDGTWIAHPGLLPTALSVEQHEMPTPNQIHVQKVEI---- 411
Qy 174 TONKLAADLEPSGDIVLK-----TIVYPRKSIATGTPKERLMFPAKA----- 217
Db 412 TEADLVDTNIPDGKLTWKGVSAIYIGLWYVESWLRGLGC-VPINNLMEDRAATAEVSRLQ 470
Qy 218 -----ADKFGKVATPLALEEFI-----AERAPTLHGLFLSCDLVKPSES-RI 259
Db 471 LYSWCKHAKVMDDTGKTITP-----EFISKLIDBEAER-----CAANKFNPKFKI 515
Qy 260 KVICMERQLD 269
Db 516 AADCLKKEIN 525

RESULT 4
T23791
hypothetical protein M176.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T23791
R:Wilkinson, J.
Submitted to the EMBL Data Library, August 1996
A:Reference number: 219799
A:Accession: T23791
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-490 <WIL>
A:Cross-references: EMBL:Z78412; PIDN:CAB01655.1; GSPDB:GN00020; CESP:M176.2
A:Experimental source: clone M176
C:Genetics:
A:Gene: CESP:M176.2
A:Map position: 2
A:Introns: 35/3; 259/2; 294/1; 471/3
C:Superfamily: glutathione synthase

Query Match 4.6%; Score 107.5; DB 2; Length 490;
Best Local Similarity 22.7%; Pred. No. 1.1;
Matches 109; Conservative 50; Mismatches 149; Indels 173; Gaps 29;

Qy 9 TLLPKPFP--YVLSQALNLSNKDHTKWWYSTAPMPATMAGAGYDVHAQYKFLC-IHREVI 65
Db 59 TLLPSPFPKVVFEAVHIQN-----LFASLYHFIAY-----EFDELIDIHKNVV 102
Qy 66 -----IPALGPVPEKG--QPM-----HWKSHLTRFRGLPFELSN----- 97
Db 103 KTDDFTRNWVEILKKVKAQGLKQPVLTATQRSYMCHEKQYSAEYGLK-QEILNNTASSM 161
Qy 98 --YSKSLLRFAFPEPLGSLTGTGKDDPPNTQAIRPVQLDKAMVPGDLLEWFDHFTK---- 151

Db 162 GAHALRLTEWHIRVLKAL-NISDDVI--QRAIPENKPIPMIAEALPKAW-SHFSNPAVV 217
QY 152 LVVSESEARTLTD-ROEIRPVFK-----TO--NKLA-----ADLESPGDIVLK 191
Db 218 LVVVENVNQIDORHVEYELEKLGVPMTCIIRNLTOCYEQLSLNDRSLMDIGRQVAI 277
QY 192 TYI-----YPRIK-----SIATGPKERLMFDAKADKFGKVATPLALIEE 233
Db 278 VYFRAGYSPDHYSTKEWARERVELSTAIKTPWIGLQ--VANTYKTKQCVLSEGVLER 334
QY 234 FIAERAPTLGLHFLSCDLVKPESRIKVCYMERQLDL-ASIEGIWTLNGRRNDPPTLD-- 290
Db 335 FIG-----KPREAR-----DIRASFAGMWAL--ENTDEVTMKVV 366
QY 291 -GLDALRELWQLLPVTEG-----LCPLNCEYEPGTSPOEQLPILNFTLSP 336
Db 367 AGAQKHEAPVLKPOQTEGAALHTGDMVQMRELPE-----SEEGATILMEKLRP 417
QY 337 -----KSALPEPQIYFPAFGQND-----KTIAEGLATFPFESRGWGLAKS 376
Db 418 MIENVLVLAKKPIITPAKAVSELGVGYAFGRKDAPELKTAGHLLTKPSTAVGGVAAG 477
QY 377 Y 377
Db 478 H 478

RESULT 5

S74207

lipoxigenase (EC 1.13.11.12) - cucumber

C:Species: Cucumis sativus (cucumber)

C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000

C:Accession: S74207; S74137

R:Hoehne, M.; Nellen, A.; Schwennesen, K.; Kindl, H.

Eur. J. Biochem. 241, 6-11, 1996

A>Title: lipid body lipoxigenase characterized by protein fragmentation, cDNA sequence a

A:Reference number: S74137; MUID:97054584; PMID:8898881

A:Accession: S74207

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-878 <HOE>

A:Cross-references: EMBL:X92890; NID:G1296511; PIDN:CAA63483.1

A:Experimental source: tissue cotyledones; clone PCSLBLOX221

A:Accession: S74137

A:Molecule type: protein

A:Residues: 196,'X',198-204;218,'FX',221-223,'XX',226-228;305-309,'XX',312-316;441-460;8

A:Superfamily: lipoxigenase

C:Keywords: oxidoreductase

Query Match 4.3%; Score 101; DB 2; Length 878;
Best Local Similarity 21.7%; Pred. No. 8.5;
Matches 94; Conservative 49; Mismatches 129; Indels 162; Gaps 26;

QY 61 HREVIIPALQ-----PYPEK---GQPMHWKSHLTRGLPFELSFNYSKSLRFAFEPLGS 112
Db 235 HR-----PILGGTTEYPYPRGRTPRPSRDH-----NYESRL-----SPIMS 273
QY 113 LT--GTKDDDFN-----TQAIRVLQDKAMVPGDLWDFHFTKALVVS 155
Db 274 LDIVYFVKDENFGHLKMSDFILGYTLKALSISIKPGLQSIQIDVTP---NEFDNF----- 322
QY 156 EEEARTLLDSDIIP--VFRTQNKLAADLEPSCDIVLKYIYPRIKSIATGPKERLMED 213
Db 323 -KEVDNLFEGFFIPNFAKT--LTEDLTP-----PLFKALVNDGKFLKPP 367
QY 214 AIXAADKFGKV--ATPLALIEEPIAERAPTLGLHFLSCDLVKPSESRIKVCYMERQLDLA 271
Db 368 TPEVV-KDNKIGSTDEEFAREMLAGNPPLLI-----RRLEAFPTSKLDP- 412
QY 272 SIEGIWTLNGRRNDPPT-----LDGLDALRELWQ-----LLPVTGELCLPLNCFYEPGPT 320
Db 413 -----NVYGNQNSTITEEHKIGLDGLTVDEAMKQNRILYIVDFHDALMPLYLTRMNATST 466

QY 321 SPQQLPFII---NFTLSP---KSALPEPO-----IYFPAFGQNDKTIASGLATFF 365
Db 467 KTYATRLLKDDGTGTLKPLVIELALPHPODQGLGALSKLYFFPAENGVOQKSI----- 518
QY 366 ESRGWGLAKSYPADLASYYPDVLDQTANHLQAWISFYSGKKPKFMSVYLHT----FFAFS 422
Db 519 ---W-QLAKAY-----VTVNDVGYHQLISHWLHTHAVLEPFV 551
QY 423 AAA-QEVAMCHDGH 435
Db 552 IATHRQLSVLPHIH 565

RESULT 6

A82255

hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: A82255

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.

R:Heidelberg, J.F.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, I.

J.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: A82255

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1621 <HEI>

A:Cross-references: GB:AEC04181; GB:AEC03852; NID:9655454; PIDN:AAF94159.1; GSPDB:GN00:

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0998

A:Map position: 1

Query Match 4.3%; Score 101; DB 2; Length 1621;
Best Local Similarity 20.5%; Pred. No. 21;
Matches 67; Conservative 47; Mismatches 99; Indels 114; Gaps 16;

QY 116 TKDDPFTQAIRPVLQDKAMVPGDLWDFHFTKALVSEEEARTLLDRDIEIPVKTQ 175
Db 1295 TEEDALADAQLEPAE--SEVEPELEJ-----ASVLEEDSFTELD-ELDIPYTEK 1343
QY 176 NKLA-ADLEPSGDIVLKYIYPRIKSIATGPKERLMFDAKADKFGKVATPLALIEEF 234
Db 1344 DALADAQLEPAESEVPELGDDETETLAQETESDALVAD-----EDLLASVESA 1392
QY 235 IAEARPTLLGHFLSCDLVKPSESRI-----KVCMERQLDLASIE 274
Db 1393 VDEVOPELLG---ATQVVPPTQSLANKAPDEEALHDLSDNPQGEKPFSPDRPLDAKTI- 1448
QY 275 GIWTLNGRRNDPPTLDGLDALREL-----WQLLPVTGELCLPLNCFYBFGTSPQQLPFII 330
Db 1449 -----DSAGND-IDAMLQMGEDWNGFHLT-----PDQQAQLPDDV 1483
QY 331 NFTLSPK-----SALPEQIYFPAFGQNDKTIASGLATFFESRGWGLAKSYPADLAS 383
Db 1484 -----PEDEQAIWASETEPOA-----KPENNGS-----QEDLLD 1513
QY 384 YYPDVD-LQTANHLQAWISFYSGKKKP 409
Db 1514 FDPQRDGYMTIDELWAQVESEEQGLNP 1540

RESULT 7

G87379

heptosyltransferase family protein (imported) - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 30-Sep-2002

C:Accession: G87379

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A>Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A/Reference number: A87249; MUID:21173698; PMID:11259647
 A/Accession: G87379

A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-323 <STO>
 A/Cross-references: GB:AE005673; NID:gl3422349; PIDN:AAK23035.1; GSPDB:GN00148
 C/Genetics:
 A/Gene: CCI051
 C/Superfamily: ADP-heptose-LPS heptosyltransferase II

Query Match 4.3%; Score 100; DB 2; Length 323;
 Best Local Similarity 19.8%; Pred. No. 2.3;
 Matches 53; Conservative 33; Mismatches 90; Indels 92; Gaps 10;

QY 178 LADLPESGDIVLK-----TIIYRIK-SIATGTPKERL-----MPDAIFA---ADKF 221
 DB 9 LVIKLGALGDFVLALAMRKIRBAHPKAKITLLTPPEALAKLSYFNSVETDGRPPDF 68
 QY 222 GKVATPLAILEEFIAERA-----PTLLGHFLSCDLVKPSESRIKV 261
 DB 69 GQTLAMLGRIKARYDEVYDLOTNSTNGYFQMLRPPFPQWNGIAGVCCSLPQKGSARLTM 128
 QY 262 YCNERQLDLASIGIWTNGRRNDPDTLGLDALRELWQLLPVTEGLCLPLNCFPYEPGTS 321
 DB 129 HTLERQADQLRAAGIW-----PDAETPGSA 154
 QY 322 PQQLPPIINFTLSPK-----SALPEQIYFPAPGQNDKTIAGLATFFESRCG----- 370
 DB 155 PPDLNLKILKHEKPRVAGAAAPRYVLP-----VPGSAHURLEKR-WPVECVQAQ 204
 QY 371 -GGIAKSPADLASYYPDVDTLOFANHLQ 397
 DB 205 LGSLLAKGLDIVIGGPQBSAMARHIQ 232

RESULT 8
 JC4240
 lanosterol 14alpha-demethylase (EC 1.14.14.-) cytochrome P450 51 - rat
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 21-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 28-Jul-2000
 C/Accession: JC4240; PC4069
 R/Sloane, D.L.; So, O.Y.; Leung, R.; Scarafia, L.B.; Saidou, N.; Jarragin, K.; Swinney, Gene 151, 243-248, 1995
 A/Title: Cloning and functional expression of the cDNA encoding rat lanosterol 14-alpha
 A/Reference number: JC4240; MUID:95394364; PMID:7665087
 A/Accession: JC4240
 A/Molecule type: mRNA
 A/Residues: 1-486 <SLO>
 A/Cross-references: GB:U17697; NID:G699395; PIDN:AAA87074.1; PID:G699396
 A/Accession: PC4069
 A/Molecule type: protein
 A/Residues: 43-60; 206-213; 254-267; 369-387; 420-428; 433-443 <SL2>
 A/Experimental source: liver
 A/Comment: This enzyme is a cytochrome P-450 enzyme which catalyzes one of the essential
 C/Superfamily: human cytochrome P450 CYP51; cytochrome P450 homology
 C/Keywords: chromoprotein; heme; iron; liver; metalloprotein; monooxygenase; oxidoreduct
 F;251-454/Domain: cytochrome P450 homology <P45>
 F;432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 4.3%; Score 99.5; DB 2; Length 486;
 Best Local Similarity 19.3%; Pred. No. 4.7;
 Matches 92; Conservative 62; Mismatches 190; Indels 133; Gaps 19;

QY 39 MFATMAGAGYDVAQYKP-LCHRSVITPAGLPYEPK-GQPMHWKSHLTRG-LPFELS 95
 DB 9 LLSLTLLACFTLSLVFLAVGHVQVLPAGAKSPFYTSYPIPLFGLHAIAFGKPIEFL 68
 QY 96 FN-YKSLRLFAFEPLGS-----LTGTKDDPFNTQAI-----RPV----- 129
 DB 69 ENAYEKYGVFSFTMGKFTYLLGSDAAALLNSKEDLNAAEYVGRITFTVFGKWAY 128

QY 130 -----LQDLKAMVPGDLLEWFDHFTKALVGBEEAR----- 160
 DB 129 DVPNAVLEQKILKSGNIA--HFQYVSIIEKEAKYFKSWGSGSERNVFEALSELI 185
 QY 161 -----TLDRDIEIPVFKTONKLAADLE-----P8-----GDIVLKT 192
 DB 186 ILTASHCLHGKEIRSQLNEKVAQIYADLDGGFSHAAMLPGWLPFLPSFRDRRAHREIKN 245
 QY 193 YIYRI-KSIATGTPKERLMFDAIKADKFGKATP-----LAILLEFIAERAPTL 242
 DB 246 IFYKAIKRRRLSKPEADIIQLTLDSTYKDGRELTDDIAGMLIGLLAGCHTSSTSAW 305
 QY 243 LGHFLSCDLVKPSESRI---KVVC-----MERQLDLASIE-----GIWTL 279
 DB 306 MGFFLARD--KPLQDKCYLQKTVCGEDLPPLTVQELKDLNLLDRCIKETLRPPIMTM 363
 QY 280 NGRNDPDTLGLDALRELWQLLPVTEGLCLPLNCFPYEPGTSPOQLPFIINFTLSPKSA 339
 DB 364 MRMAKTPTVAG-----YTIPPGHQVCVSPVQNRKDSWVERLDFNPDRYLQNP 415
 QY 340 LPEQIYFPAPGQNDKTIAGLATFFESRCGWLAKSPADLAS-YYPDVDTLOFANH 395
 DB 416 SGEXFAYVPGAGHRCIGENFAVQIKTIWMLRIYEFDLINGYFVSVNTMIH 472

RESULT 9
 JC4758
 lanosterol 14alpha-demethylase (EC 1.14.14.-) cytochrome P450 51 - rat
 A/Alternate names: pRT-9 protein
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 28-Jul-2000
 C/Accession: JC4758; PC4170; JC5888; JC2334
 R/Aoyama, Y.; Noshiro, M.; Gotoh, O.; Imaoka, S.; Funae, Y.; Kurosawa, N.; Horiuchi, T.; J. Biochem. 119, 926-933, 1996
 A/Title: Sterol 14-demethylase P450 (P45014DM) is one of the most ancient and conserved
 A/Reference number: JC4758; MUID:56389999; PMID:8797093
 A/Accession: JC4758
 A/Molecule type: mRNA
 A/Residues: 1-503 <AOY1>
 A/Cross-references: DDBJ:DF5681; DDBJ:D29962; NID:gl020094; PIDN:BAA09529.1; PID:G870754
 A/Accession: PC4170
 A/Molecule type: protein
 A/Residues: 52-109; 305-319; 324-334; 467-482; 488-499 <AOY2>
 A/Experimental source: liver
 R/Noshiro, M.; Aoyama, Y.; Kawamoto, T.; Gotoh, O.; Horiuchi, T.; Yoshida, Y. J. Biochem. 122, 1114-1121, 1997
 A/Title: Structural and evolutionary studies on sterol 14-demethylase P450 (CYP51), the
 A/Reference number: JC5888; MUID:98159318; PMID:9498553
 A/Accession: JC5888
 A/Molecule type: DNA
 A/Residues: 1-503 <NOS>
 A/Cross-references: DDBJ:AB004087
 R/Aoyama, Y.; Funae, Y.; Noshiro, M.; Horiuchi, T.; Yoshida, Y. Biochem. Biophys. Res. Commun. 201, 1320-1326, 1994
 A/Title: Occurrence of a P450 showing high homology to yeast lanosterol 14-demethylase (

A/Reference number: JC2334; MUID:94296405; PMID:8024575
 A/Accession: JC2334
 A/Molecule type: mRNA
 A/Residues: 74-503 <AOY>
 A/Cross-references: DDBJ:D29962
 A/Experimental source: liver
 C/Genetics:
 A/Gene: CYP51
 A/Introns: 59/3; 91/3; 150/3; 193/1; 251/2; 291/2; 356/3; 388/3; 445/1
 C/Function:
 A/Description: catalyzes the removal of the 14-methyl group of 14-methylsterols
 C/Superfamily: human cytochrome P450 CYP51; cytochrome P450 homology
 C/Keywords: chromoprotein; heme; iron; liver; metalloprotein; monooxygenase; oxidoreduct
 F;308-471/Domain: cytochrome P450 homology <P45>
 F;445/Binding site: heme iron (Cys) (axial ligand) #status predicted
 Query Match 4.3%; Score 99.5; DB 2; Length 503;
 Best Local Similarity 19.3%; Pred. No. 4.9;

A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70922
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-344 <COL>
A;Cross-references: GB:295150; GB:AL123456; NID:g2250708; PID:CA080361.1; PID:g2076700
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: RV3127
C;Superfamily: Mycobacterium hypothetical protein RV3131

Query Match 4.2%; Score 97.5; DB 2; Length 344;
Best Local Similarity 20.7%; Pred. No. 4.1;
Matches 65; Conservative 38; Mismatches 94; Indels 117; Gaps 14;

QY 80 HKSHLTFGLPELSFNYSKSLRFAFBLGSLTGK-----DDPFN 122
DB 73 HWQANITRFPPQ-----NQPDQATVEFSPIDHVTAGQNRQAQILQRTDLRPFDSMY 127

QY 123 TQAIRPVLOD-LKAMPVGLDLEWDFHFTKALVSE----- 156
DB 128 WHLFEPAIRDVADKVAQLDVSDDQRTSLVVASQLSEVLRRDPYHAELEWWTSPFVL 187

QY 157 -----EARTLLDRDIEIPVFTONKLAADLEPSGDIVLKYIIPRIKSIAT 203
DB 188 AHGVPPDTLASDAERLURVDLGRD--PFRVSQNRRA---ELADD-----RSKVLVL 233

QY 204 GTPKERMFDIAKAADKFGKATPLAILBEFIAERAPTLILGHFLSCDLVKPSESRIKYVC 263
DB 234 STPSS-TRADALRC-----GVLSTI-LLECTMAGMATCTLT-----LISSDSRDIRVG 282

QY 264 MERQDLASIEGTWLNGRNDETDGLDALRELWQLPVTBGLCPNCFYEGTSPQ 323
DB 283 LTRQ-----RGEPOAL-----IRVGTIA--PPLAAVPAFPTPR 311

QY 324 EQLPFIINFTLSPK 337
DB 312 RPLDSVLQIRQTPB 325

RESULT 13
A71221
aldehyde-ferredoxin oxidoreductase (BC 1.-.-.-) PH0028 [imported] - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 03-Nov-2000
C;Accession: A71221
R;Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: A71221
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-621 <RAW>
A;Cross-references: GB:AP000001; NID:g23236128; PID:BAA29096.1; PID:g3256413
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0028
C;Superfamily: probable aldehyde ferredoxin oxidoreductase aor-4
C;Keywords: oxidoreductase

DB 102 ----FLKGFHFDALIEGKSEEPVLYIHDGVDILPAGELW---GKGN-----YETA 147
QY 296 RELHQ-----LLPVTGCLPLNCFY-----BEGTS 321
DB 148 RELMKYPEASIASIGPAGERLVRIANIYDTPASRGGLGAVMGSKLKAIVVEGK 207
QY 322 PQEQLP-----FINTFLSPKALPEFQIYFPFAGQNDKTIAREGLATFFESRGWG- 371
DB 208 PEVANPEFEALWNEFYERFSTDPK-----YEHSRYGT 241
QY 372 -----GLAKSY-----PADLASYPDVDLQTNHLQAW-----I 400
DB 242 TDGLRSSASLGMSPAYNFSRPYIPEELASKLAGDEVKYEVEPEWYIHGKSCPIKARYI 301
QY 401 SFSYKGGK 408
DB 302 EVEYKGRK 309

RESULT 14
S69079
hypothetical protein YPR097w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein P9513.1
C;Species: Saccharomyces cerevisiae
C;Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 19-Apr-2002
C;Accession: S69079
R;Couch, J.
A;Description: The sequence of S. cerevisiae cosmid 9513.
A;Reference number: S69057
A;Accession: S69079
A;Molecule type: DNA
A;Residues: 1-1073 <COU>
A;Cross-references: EMBL:U51033; NID:g1230676; PID:g1230699; GSPDB:GN00016; MIPS:YPR097w
C;Genetics:
A;Gene: MIPS:YPR097w
A;Cross-references: SGD:S0006301
A;Map position: 16R

Query Match 4.1%; Score 95.5; DB 2; Length 1073;
Best Local Similarity 19.5%; Pred. No. 32;
Matches 66; Conservative 61; Mismatches 116; Indels 95; Gaps 17;

QY 9 TLLPKPYYVLSQLNLSNKDHTK-----WW--YSTAPMFATMMAG-AGYDVH 52
DB 556 SLLKDEKYLMSLVEI--KNTKVEDLSPLLQDFVWCKYIISMTYQMFGLNDNSYEL 613

QY 53 AQYKFLCIHREVIIPALGPYPERGQPMHWSHITRFLPELPSFNYSKSLRFAFPPLGS 112
DB 614 TQIRRL--HK-----LMPYVMGQIMKFTNPIAINRGMIEL-----FMAQPF 655

QY 113 LTGTDPPFTQAIRPVLODKAVVGLDLEWDFHETKALVWSEEBARTLLDRDILP 172
DB 656 -----HSLLOTFSTILTDDLK-----TQVAIKE-----LEKI----- 695

QY 173 KTONKLAADLEPSGDIV--LKYIYIPRIKSIATGTPKRLMP--DAIKAADKFGKVP 227
DB 686 -----AENDPGASVVTKLDFVF-----NNDTKDEHDTKLTMDAVNAESSENNMPVP 734

QY 228 LAILEFIA-----ERAPTLIGHFLSCDLVKPSESRIKYCMERO-LDLASIEGTWLN 280
DB 735 LIVLMKSAANLIPDEVVAGLIESYSWKLQKEDTDALANTSEDQSGIYFTHVKDLWQLY 794

QY 281 GRNDDPETGLDALRELWOLLPVTEGLCPNCFYEP 318
DB 795 IKEHDQ-----LMRQLAQDDELQMLKAIVTMIYEP 826

RESULT 15
GNVTR
genome polyprotein 2 - tomato ringspot virus (strain raspberry)
N;Contains: coat protein

C:\Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jun-2000

C:\Accession: JQ1093
R:\Rott, M.E.; Tremaine, J.H.; Rochon, D.M.
J. Gen. Virol. 72, 1505-1514, 1991
A:\Title: Nucleotide sequence of tomato ringspot virus RNA-2.
A:\Reference number: JQ1093; NUID:91311402; PMID:1856689
A:\Accession: JQ1093
A:\Molecule type: genomic RNA
A:\Residues: 1-1882 <ROT>
A:\Cross-references: GB:D12477; GB:D01129; NID:G222674; PIDN:BA02043.1; PID:G222675
A:\Note: it is uncertain whether Met-1 or Met-122 is the initiator

C:\Genetics:
A:\Map position: segment 2
C:\Superfamily: tomato ringspot virus genome polypeptide
C:\Keywords: coat protein; glycoprotein; polypeptide
F:\321-1882/Product: coat protein #status predicted <MAT>
F:\269,295,1183,1316,1543,1561,1735/Binding site: carbohydrate (Asn) (covalent) #status P

Query Match 4.1%; Score 95.5; DB 1; Length 1882;
Best Local Similarity 20.3%; Pred. No. 72;
Matches 84; Conservative 50; Mismatches 129; Indels 151; Gaps 24;
QY 33 WYSTAMFATMAGAGYVHAQKFLC-----IHREVIIP-----ALGPYPE 74
DB 608 WSSPLFLFANFKVNRG-----ACFLQVLPQVRVLPDECMDLLSLFEDQLPEGP LPS 658
QY 75 KGQPMHWKSHLTRFGLPFELSFNYSK-----SLLRFAFEPLGSLTG 115
DB 659 ---FSWSS-----PLPLFASFKVNRGACFLQVLPARKVVSDFMDVLPFLFSP L--VSH 707
QY 116 TKDDPNTQAIRPVLODL-----KAVPGLDLE-WPDHFTKALVSESEARTLLDRDIEIP 170
DB 708 QEEPEMVPVLEAADSVDGITEAFEDDLECSFYDSYS-----DEEEAEW-----AEVP 757
QY 171 VFKTOKLAADLEPSGDIVLKTIVYPRIKSIATGTPKERLMFDAIKADKFGKVATPLAI 230
DB 758 RCKTMBELCASLTLAGD-----AEGLRKSHGVF-----LKLVTYLOS 795
QY 231 LEE-FIAERAPTLLGHFLSCDLVKP-----SESRIKVCMERQLD-----LASIEGIWTL 279
DB 796 FEESPLYSSRA-----FYSVK-VKPVYRPPKFECHIDCTCLDGNMGCEWRESVDAMWRC 848
QY 280 NGR-----RNDPETHDGL-----DALRELWQLLPVTE-----GLCPLPN 313
DB 849 PGRLLNTRKFTTDDWERYOYLRIGFNEGRYRNWRVNLNLEMDLSLHVEYPEISSAPVOS 908
QY 314 CFY-----EPGTSPEQLPFIINFTLSPKSGALPEPQIYFPFAGQNDKXTIABGLAT 363
DB 909 SLFSRVVDRCATLASSIPFVTRNC--QSSLGTP-----GLNVHTIHCEAPT 953

Search completed: July 22, 2004, 18:46:29
Job time : 21 secs

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:34:31 ; Search time 13 Seconds
(without alignments)
1750.358 Million cell updates/sec

Title: US-10-099-704-2
Perfect score: 2322
Sequence: 1 M5IKKAATLLPKPFVLSQ.....FEAFSAQAQEVAMCHDGHNP 437

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	111.5	4.8	378	1 MUTY_HAEIN	P44320 haemophilus
2	111.5	4.8	555	1 MASY_PICAN	P21360 pichia angu
3	110.5	4.8	462	1 ACDS_METKA	Q8txf1 methanopyru
4	99.5	4.3	503	1 CPSI_RAT	Q64654 rattus norv
5	98.5	4.2	2649	1 P285_HUMAN	Q9byk8 homo sapien
6	95.5	4.1	1882	1 POL2_TRSVR	P25247 tomato ring
7	95	4.1	610	1 IRBP_MOUSE	P49194 mus musculu
8	94.5	4.1	430	1 SYH_CHLPN	Q9z7p1 chlamydia p
9	94.5	4.1	896	1 POL_HILLA	P03362 human t-cel
10	94	4.0	910	1 SYV_SYNY3	Q55522 synechocyst
11	93	4.0	833	1 SYL_STRPN	Q978s0 streptococ
12	93	4.0	2492	1 POLN_EEVT	P27282 venezuelan
13	92.5	4.0	862	1 ADHE_CLOAB	P33744 clostridium
14	92	4.0	282	1 YMA8_CABEL	P34453 caenorhabdi
15	92	4.0	443	1 HSLU_HELPJ	Q9zlw1 helicobacte
16	92	4.0	500	1 C72W_ARATH	Q91lp5 arabidopsis
17	92	4.0	1017	1 GCSP_SCHPO	Q09785 schizosacch
18	92	4.0	2485	1 POLN_EEVV3	P36327 venezuelan
19	91.5	3.9	466	1 NMT_SCHPO	O43010 schizosacch
20	91.5	3.9	986	1 DPOL_NPVBM	P41712 bombyx mori
21	91	3.9	609	1 TYDP_MOUSE	Q8bj37 mus musculu
22	91	3.9	766	1 HIFA_ONCMY	Q985w2 oncorhynchu
23	91	3.9	833	1 SYL_STRR6	Q8dbt6 streptococ
24	90.5	3.9	521	1 SYL_METJA	Q58477 methanococ
25	90.5	3.9	833	1 DPOI_THEFI	O52225 thermus fil
26	90.5	3.9	852	1 RA54_SCHPO	P41410 schizosacch
27	90.5	3.9	2278	1 FAB1_YEAST	P34756 saccharomyc
28	90	3.9	862	1 SYV_THETH	P96142 thermus the
29	90	3.9	2492	1 POLN_EEVV3	P36328 venezuelan
30	90	3.9	4303	1 PKDI_HUMAN	P98161 homo sapien
31	89	3.8	765	1 YK07_YEAST	P36122 saccharomyc
32	89	3.8	985	1 AGLU_ASPOR	Q12558 aspergillus
33	88.5	3.8	451	1 KPT3_MOUSE	Q04899 mus musculu

34	88.5	3.8	554	1 PIGS_HUMAN	Q36552 homo sapien
35	88.5	3.8	791	1 SYFB_XANCP	Q8p726 xanthomonas
36	88.5	3.8	1277	1 PDS5_YEAST	Q04264 saccharomyc
37	88.5	3.8	1356	1 Y195_HUMAN	Q12767 homo sapien
38	88.5	3.8	1420	1 SRB9_YEAST	P38931 saccharomyc
39	88	3.8	504	1 CP3B_MOUSE	Q64459 mus musculu
40	88	3.8	761	1 METE_AQUAE	O57606 aquifex aeo
41	88	3.8	2763	1 TEGU_VZVD	P09278 varicella-z
42	87.5	3.8	488	1 ACDS_METJA	Q37576 methanococc
43	87.5	3.8	680	1 OPDA_SALTY	P27237 salmonella
44	87.5	3.8	831	1 PPOF_THETH	P30313 thermus the
45	87.5	3.8	1520	1 GLTB_EAGSU	P39812 bacillus su

ALIGNMENTS

RESULT 1
MUTY_HAEIN
ID MUTY_HAEIN STANDARD; PRT; 378 AA.
AC P44320;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE A/G-specific adenine glycosylase (EC 3.2.2.-).
MUTY OR HI0759.
OS Haemophilus influenzae.
GN Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellales; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geochagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: Adenine glycosylase active on G-A and C-A mispairs (By similarity).
CC -!- COPACTOR: Binds a 4Fe-4S cluster which is not important for the catalytic activity, but which is probably involved in the proper positioning of the enzyme along the DNA strand (By similarity).
CC -!- SIMILARITY: Belongs to the nth/mutY family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32760; AAC22418.1; -.
CC PIR; C64091; C64091.
CC HSRP; P17802; IMUY.
CC TIGR; HI0759; -.
CC InterPro; IPR003265; Endo_3c.
CC InterPro; IPR004035; EndoIII_FCL.
CC InterPro; IPR004036; EndoIII_HhH.
CC InterPro; IPR003651; FeS_bind.
CC InterPro; IPR000445; HHH.
CC InterPro; IPR005760; MutY.
CC Pfam; PF00730; HHH-GPD; 1.
CC Pfam; PF00633; HHH; 1.

DR SMART; SM00478; ENDO3c; 1.
 DR SMART; SM00525; FES; 1.
 DR TIGRFAMS; TIGR01084; muty; 1.
 DR PROSITE; PS00784; ENDONUCLEASE_III_1; 1.
 DR PROSITE; PS01155; ENDONUCLEASE_III_2; 1.
 KW DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;
 Complete proteome.
 FT METAL 197 197 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 204 204 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 207 207 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 213 213 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 378 AA; 43505 MW; C443F625131B2A21 CRC64;
 Query Match 4.8%; Score 111.5; DB 1; Length 378;
 Best Local Similarity 22.8%; Pred. No. 0.14;
 Matches 79; Conservative 48; Mismatches 127; Indels 93; Gaps 21;
 QY 142 LEWFDHTK-----ALVSVSEEAATLLDRDIEIPV-----KTQNKLA 180
 DB 16 LAWYDFGRKHLFPWQNKTLYGWLVSEVLMQQTQVATV-----IPYERFIKTFPNITA 69
 QY 181 DLEPSPGDIVLTKY-----IYPRIKSIATGTGPKER-----LMFDAKAADKFGKVATPL 228
 DB 70 LANASQDEVLLHWLTGLGYARARNLHKAQKVRDEPNFGNPTNFQVWALSGVGR-STAG 128
 QY 229 AILBEETIAERAPFLHGHFSLCDLVKPSER---IKVYCMERQDLASTIEGIWTLNGRRND 285
 DB 129 AILLSVINOPYPLIDGN-----VRLVARYFAVEGWSGEKKVE---NRLWALT-EQVT 177
 QY 286 PET--LDGLDALRELWQLLPVTE---GLCPLP-NCFYEPGTSPOBOLFFIINFILSPKS 338
 DB 178 PTRVADFNQAMMDIGAMVCMRTKPCDLCPLNIDCLAYKVN-WEKFP-----AKPKK 231
 QY 339 ALPEPQIYPPAFQNDKTIAGLAEIATFFESRG-WGLAKSYPP--ADLASYPYDVIDQTANH 395
 DB 232 AMPEKTYTFLILSKNGKVCLEQR---ENSLWGLGF-CFPQFEDKSSLHLLEAQBKVTH 286
 QY 396 LQAWISPSYKPKPYMSVLVHFEAPSA-----AAQEVAMCHDGN 436
 DB 287 YQWPSFR-----HTFSHPHLDIHPYAEEMESTLVCVQAN 321
 RESULT 2
 ID MASY_PICAN STANDARD; PRT; 555 AA.
 AC P21360;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Malate synthase, glyoxysomal (EC 2.3.3.9).
 GN MAS.
 OS Pichia angusta (Yeast) (Hansenula polymorpha).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4905;
 RN [1]
 RC STRAIN=MAYA CB54732;
 RP MEDLINE=90273778; PubMed=2349836;
 RX Bruinenberg P.G., Blaauw M., Kazemier B., Ab G.;
 RA "Cloning and sequencing of the malate synthase gene from Hansenula
 RT polymorpha";
 RL Yeast 6:245-254 (1990).
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CoA.
 CC -1- PATHWAY: Glyoxylate bypass; second step.
 CC -1- SUBCELLULAR LOCATION: Glyoxysomal.
 CC -1- INDUCTION: By ethanol.
 CC -1- SIMILARITY: Belongs to the malate synthase family.
 DR PIR; S09294; SYHQA.
 DR InterPro; IPR001465; Malate synthase.
 DR InterPro; IPR001465; Malate synthase.

DR TIGRFAMS; TIGR01344; malate syn A; 1.
 DR PROSITE; PS00510; MALATE SYNTHASE; 1.
 KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle; Glyoxysome.
 SQ SEQUENCE 555 AA; 63255 MW; 98137B7B2E3E303 CRC64;
 Query Match 4.8%; Score 111.5; DB 1; Length 555;
 Best Local Similarity 23.5%; Pred. No. 0.24;
 Matches 73; Conservative 39; Mismatches 89; Indels 109; Gaps 20;
 QY 16 YVLSQALNLNKHDT-----KWYSTAPMFATMAGAGYDVHAQYKFLCIHREVIIPAL 69
 DB 269 YQLEDEV-YALRDHSAGLNGRWY---NFSII-----XRLNQKHLP-- 309
 QY 70 GPYPEKQPMHWSHLTRFGLPFELSFNYSKSLRFAFE---PLGSLTGT---KDDP-F 121
 DB 310 -----DRHQVTMTVPF--MTNVYKLIKIKCHKRGVHANGMAATIPKDDPEK 355
 QY 122 NTQAIRPVLOD-LKAMVPGLDLEWFDH---FTKALVSVSEEAET---LLDRDIEIPVK 173
 DB 356 NAAAMEAVROCKLREVLAGHDGTWIAHPGLLPALTALSVFQEHMPTPNQIHVQKNVEI--- 411
 QY 174 TQNKLAADLPSPGDIVLK-----TYIYPRIKSIATGTGPKERIMFDAIKA----- 217
 DB 412 TEADLVDTNIPDGKITMKGVSANIYIGLNYMESWLRGLGC-VPINLMEDAATAEVSRLQ 470
 QY 218 -----ADKFGKVATPLALEBFI-----ABRAPTLGHFLSCDLVPSSES-RI 259
 DB 471 LYSCKHAKVMDDTGKTIPT-----BFISKLIDEEAER-----CAANKPNKFKI 515
 QY 260 KVMCMEROLD 269
 DB 516 AADCLKKEIN 525
 RESULT 3
 ID ACDG_METKA STANDARD; PRT; 462 AA.
 AC O8TXE1;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Acetyl-CoA decarboxylase/synthase complex gamma subunit (EC 2.1.1.-)
 DE (ACDS complex gamma subunit) (ACDS complex methyltransferase)
 DE (Corrinoid/iron-sulfur component large subunit).
 GN CDHE OR MK0723.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 OC Methanopyrus.
 OX NCBI_TaxID=2320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezhevaeva K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozhavkin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 RT and monophyly of archaeal methanogens";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 CC -1- FUNCTION: Part of a complex that catalyzes the reversible cleavage
 CC of acetyl-CoA, allowing autotrophic growth from CO(2) (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: Co(1) + CH(3)-H(4)SPT + H(1) = Co(III)-CH(3) +
 H(4)SPT.
 CC -1- COFACTOR: Binds 1 4Fe-4S cluster and factor III (Probable).
 CC -1- SUBUNIT: Heterodimer of delta and gamma chains. The ACDS complex
 CC is made up of alpha, epsilon, beta, gamma and delta chains with a
 CC probable stoichiometry of (alpha(2)epsilon(2)) (4)-beta(8)-
 CC (gamma(1)delta(1)) (8) (Potential).
 CC -1- SIMILARITY: The iron-sulfur centers are similar to those of
 CC bacterial-type 4Fe-4S ferredoxins

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE010365; AA001937.1; -.	
HMAP; MF_011136; -, 1.	
InterPro; IPR004486; CdhD.	
InterPro; IPR007202; FeS.	
Pfam; PF03599; CdhD; 1.	
Pfam; PF04060; FeS; 1.	
PROSITE; PS00198; 4FE4S FERREDOXIN; FALSE NEG.	
Transferase; Methyltransferase; Metal-binding; Cobalt; Iron;	
Iron-sulfur; 4Fe-4S; Complete proteome.	
METAL 18 18 IRON-SULFUR (4FE-4S) (PROBABLE).	
METAL 21 21 IRON-SULFUR (4FE-4S) (PROBABLE).	
METAL 26 26 IRON-SULFUR (4FE-4S) (PROBABLE).	
METAL 43 43 IRON-SULFUR (4FE-4S) (PROBABLE).	
SEQUENCE 462 AA; 50852 MW; 1E97B54407368D3B CRC64;	

Query Match 4.8%; Score 110.5; DB 1; Length 462;
Best Local Similarity 27.0%; Pred. No. 0.22;
Matches 62; Conservative 28; Mismatches 91; Indels 49; Gaps 12;

[illegible]

RESULT 4	CP51_RAT	STANDARD;	PRT;	503 AA.
ID	CP51_RAT			
AC	Q64654;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Cytochrome P450 51A1 (EC 1.14.13.70) (CYP1I)			
DE	demethylase) (lanosterol 14-alpha demethylase)			
GN	CYP51A1 OR CYP51.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RC	MEDLINE=94236405; PubMed=8024575;			
RA	Aoyama Y., Funae Y., Noshiro M., Horiuchi T., Yoshida Y.;			
RT	"Occurrence of a P450 showing high homology to yeast lanosterol 14-			
RT	demethylase (P450 14DM) in the rat liver.";			
RL	Biochem. Biophys. Res. Commun. 201:1320-1326(1994).			

RN	
KN	
I2J	SEQUENCE FROM N.A.
RP	STRAIN=Wistar;
RC	MEDLINE=981583;8; PubMed=9498553;
RX	Noshiro M., Aoyama Y., Kawamoto T., Gotoh O., Horiuchi T., Yoshida Y. ;
RA	"structural and evolutionary studies on sterol 14-demethylase P450
RT	(CYP51), the most conserved p450 monooxygenase: I. Structural analyses
RT	
RT	

of the gene and multiple sizes of mRNA." ;
J. Biochem. 122:1114-1121(1997).
(3)
SEQUENCE OF 18-503 FROM N.A.
TISSUE=Liver;
MEDLINE=95394364; PubMed=7665087;
Sloane D.L., So O.Y., Leung R., Scarafia L.E., Saldou N., Jarnagin K.,
Swinney D.C.;
"Cloning and functional expression of the cDNA encoding rat
lanosterol 14-alpha demethylase." ;
Gene 161:243-248(1995).
C-!- FUNCTION: Catalyzes C14-demethylation of lanosterol; it transforms
lanosterol into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol.
-!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
oxidized flavoprotein + H(2)O.
-!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
NADP(+) + 3 H(2)O.
-!- PATHWAY: Cholesterol biosynthesis.
-!- SUBCELLULAR LOCATION: Microsomal (Potential).
-!- SIMILARITY: Belongs to the cytochrome P450 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL;	DS5681;	BAA09529.1;	-	
EMBL;	AB004096;	BAA20354.1;	-	
DR	EMBL;	AB004087;	BAA20354.1;	JOINED.
DR	EMBL;	AB004088;	BAA20354.1;	JOINED.
DR	EMBL;	AB004089;	BAA20354.1;	JOINED.
DR	EMBL;	AB004090;	BAA20354.1;	JOINED.
DR	EMBL;	AB004091;	BAA20354.1;	JOINED.
DR	EMBL;	AB004092;	BAA20354.1;	JOINED.
DR	EMBL;	AB004093;	BAA20354.1;	JOINED.
DR	EMBL;	AB004094;	BAA20354.1;	JOINED.
DR	EMBL;	AB004095;	BAA20354.1;	JOINED.
DR	EMBL;	U17697;	AAA87074.1;	-
DR	PIR;	JC4240;	JC4240.	
DR	HSSP;	P77901;	1E9X.	
DR	InterPro;	IPR001128;	Cytochrome_P450.	
DR	Tram;	PF00067;	p450; 1.	
DR	PRINTS;	PR00385;	P450.	
DR	PROSITE;	PS00086;	Cytochrome_P450; 1.	
KW	Oxidoreductase;	Monooxygenase;	Electron transport; Transmembrane;	
KW	Heme;	Cholesterol biosynthesis;	NADP.	
FT	TRANSEM	24	44	POTENTIAL.
FT	METAL	449	449	IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT	CONFLICT	181	181	E -> K (IN REF. 3).
FT	SEQUENCE	503 AA;	56706 MW;	3398F345FF92Cf21 CRC64;
SO				

Query Match	4.3%	Score 99.5;	DB 1;	Length 503;
Best Local Similarity	19.3%	Pred. No. 2;		
Matches 92;	Conservative 62;	Mismatches 190;	Indels 133;	Gaps 19;

[illegible]

Db 203 ILTASHCLHKEIRTSQNEKVAQYADLDGGFSHAALLPGWLPFSFRRRRRAHREIKN 262
 QY 193 YIYPRI-KSIATGTPKRLMFDIAKAADKFGKVATP-----LAILEEFAIERAPT 242
 Db 263 IFYKAIGKRRLSKPEADIIQTLDSYKDGRLTDEIAGMLIGLLAGQHTSSYTSAW 322
 QY 243 LGHFLSCDLVKPESRI---KYVC-----MERQLDLASTE-----GIWTL 279
 Db 323 MGFFLARD--KPLQDKCYLEQKTVCGEDLPPLYEQKDLNLDRCIKETLRLRPPIMTM 380
 QY 280 NGRNDPDTLGDALRELWQLLVTEGCLPFCNCFYEPGTSPQEQLPFIINFTLSPKSA 339
 Db 381 MRMAKTPQTVAG-----YTTPGHQVCVPTVQRLKDSNVERLDNPDRLQDNFA 432
 QY 340 LPPEQIYFPAGQNDKTIABGLATFFESRGWGLAKSYPADLAS-YPDVIDLQTANH 395
 Db 433 SGKFAFYVPFGAGRHRCIGENFAVQIKTIWMTMLRYEFDLNGYFSPVNTYTTMIH 489

RESULT 5

P285 HUMAN
 ID P285_HUMAN STANDARD; PRT: 2649 AA.
 AC Q9BYK8; Q9BYK8; Q9BYK8; Q9C094;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Peroxisomal proliferator-activated receptor A interacting complex
 GN P1C285 OR KIA11765.
 OS Homo sapiens. (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=22199868; PubMed=12189208;
 RA Surapreddi S., Yu S., Bu H., Hashimoto T., Yeldandi A.V.,
 RA Kashireddy P., Cherkaoui-Malki M., Qi C., Zhu Y.-J., Rao M.S.,
 RA Reddy J.K.;
 RT "Identification of a transcriptionally active peroxisome proliferator-
 RT activated receptor alpha-interacting cofactor complex in rat liver and
 RT characterization of P1C285 as a coactivator.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:11836-11841 (2002).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Brain;
 RC MEDLINE=21082932; PubMed=11214970;
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RL for large proteins in vitro.";
 RL DNA Res. 7:347-355 (2000).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yanakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones.";
 RL DNA Res. 9:99-106 (2002).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS SER-788; HIS-1123;
 RP PRO-2016 AND GLN-2049.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
 RA

RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leharshah M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McKusick A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871 (2001).
 RN [5]
 RP SEQUENCE OF 1526-2649 FROM N.A., AND VARIANTS PRO-2016 AND GLN-2049.
 RC TISSUE=Spleen;
 RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;
 RT "The nucleotide sequence of a long cDNA clone isolated from human
 RL spleen.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 2204-2649 FROM N.A.
 RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuna M.,
 RA Murakawa K., Kurehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
 RA Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K.,
 RA Isogai T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Helicase that acts as a transcriptional co-activator for
 CC peroxisome proliferator activated receptor alpha (PPAR-alpha).
 CC -!- SUBUNIT: Interacts with PPARA, ESRI, THRB and RXRA. It is probably
 CC part of the peroxisome proliferator activated receptor alpha
 CC interacting complex (PRIC).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=3;
 CC IsoId=Q9BYK8-1; Sequence=Displayed;
 CC Note=No experimental confirmation available;
 CC Name=2;
 CC IsoId=Q9BYK8-2; Sequence=VSP_007297, VSP_007298;
 CC Name=1;
 CC IsoId=Q9BYK8-3; Sequence=VSP_007299, VSP_007300, VSP_007301;
 CC -!- TISSUE SPECIFICITY: Expressed in various tissues including
 CC skeletal muscle, colon, spleen, liver, kidney, lung, peripheral
 CC blood, and placenta.
 CC -!- DOMAIN: Contains 5 Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC -!- SIMILARITY: Belongs to the DNAB/NAM7 helicase family.
 CC -!- CAUTION: Ref.1 experiments have been carried out partly in rat and
 CC partly in human.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF517673; AM74197.1; -
 CC EMBL; AB051556; BA21860.2; -
 CC EMBL; AL121829; CAC15328.3; -

EMBL; AK055611; BAB70969.1; ALT_INIT.
DR InterPro; IPR001900; Ribonuclease_II.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00773; RNB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS01157; ZINC_FINGER_C2H2_2; FALSE NEG.
DR PROSITE; PS01175; RIBONUCLEASE_II; UNKNOWN_1.
KW Transcription regulation; Activator; Nuclear protein; ATP-binding;
KW Helicase; DNA-binding; Metal-binding; Zinc; Repeat; Zinc-finger;
KW Alternative splicing; Polymorphism.
DR ZN_FING 40 64 C2H2-TYPE (ATYPICAL).
FT NP_BIND 544 551 ATP (POTENTIAL).
FT SITE 667 670 DEAA BOX.
FT NP_BIND 2174 2181 ATP (POTENTIAL).
FT DOMAIN 968 1039 ALA-RICH.
FT SITE 1075 1079 LXXLL MOTIF 1.
FT SITE 1118 1122 LXXLL MOTIF 2.
FT SITE 1173 1177 LXXLL MOTIF 3.
FT SITE 2012 2016 LXXLL MOTIF 4.
FT SITE 2229 2233 LXXLL MOTIF 5.
FT VARSPLIC 1 569 Missing (in isoform 2).
FT VARSPLIC /FTID=VSP_007297.
FT VARSPLIC LICHTN -> MSSGPSR (in isoform 2).
FT VARSPLIC /FTID=VSP_007298.
FT VARSPLIC Missing (in isoform 1).
FT VARSPLIC /FTID=VSP_007299.
FT VARSPLIC Missing (in isoform 1).
FT VARSPLIC /FTID=VSP_007300.
FT VARSPLIC Missing (in isoform 1).
FT VARSPLIC /FTID=VSP_007301.
FT VARIANT N -> S (in dbSNP:438363).
FT VARIANT /FTID=VAR_015597.
FT VARIANT R -> H (in dbSNP:310632).
FT VARIANT /FTID=VAR_015598.
FT VARIANT L -> P (in dbSNP:3810485).
FT VARIANT /FTID=VAR_015599.
FT VARIANT E -> Q (in dbSNP:3810483).
FT VARIANT /FTID=VAR_015600.
FT SEQUENCE 2649 AA; 294711 MW; 841892F5521FF459 CRC64;
Query Match 4.2%; Score 98.5; DB 1; Length 2649;
Best Local Similarity 20.3%; Pred. No. 25;
Matches 70; Conservative 34; Mismatches 100; Indels 141; Gaps 16;
QY 138 PGLDLEWDFHFTKALVVS-----EEEARLLDRDIEIPVFKTQNKLA 180
DB 321 PALALE-----FNSSVASGIPSTNYRQRMHQFLYEBA-----AQQLVA 362
QY 181 DLPSGDIVLKYIYPRKSIATGTPKRLMFDPAIKADKFGKVPATPLAILFEFTAERAP 240
DB 363 KLTIRGQVFLKT-----ALQTPALNMLFAPGAL--YAEVVPVSSIMFD--TDQG 408
QY 241 TLIGHFLSCDLVKP---SESRIKVCWEROOLDASIEGIWTLNGRRNDPETLDGLDALRE 297
DB 409 FLIGRAVSTALVAPVAPDNTVEVLRR--ASSE-----QA 444
QY 298 LWOLLPVTEGLCPKNCYBPGTSPOQLFFINFILSPKS-----ALPEQIYFP 348
DB 445 LWLLLP-----RCCLALGLQPEARLVLEQFQIDPMTFRLMHAQVDTLPEEQLVVP 496
QY 349 AF-----GQNDXTIAGLATFFESRGWG-----GLAKS 376
DB 497 DLPTCALPRWSPVPLRGRNKGDELAVALIA-----GWPGDGRVPPLIIVGPFGTGKT 551
QY 377 YPADLASY-----YPDVDLOTANHLOWISFYSGKPKYMSYLHT 417
DB 552 YTLAVASLEVIRRPETKVLCTHNSAADI-----YIREYFHS 589

RESULT 6
POL2 TRSVR
ID -POL2 TRSVR
AC P25247; STANDARD; PRT; 1882 AA.

01-MAY-1992 (Rel. 22, Created)
01-MAY-1992 (Rel. 22, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
DB RNA2 polyprotein (207 kDa protein) [Contains: Coat protein].
OS Tomato ringspot virus (isolate raspberry) (TomRSV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus.
OX NCBI_TaxID=12281;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91311402; PubMed=1855689;
RA Rott M.E., Tremaine J.H., Rochon D.M.;
RT "Nucleotide sequence of tomato ringspot virus RNA-2."; J. Gen. Virol. 72:1505-1514 (1991).
RL J. Gen. Virol. 72:1505-1514 (1991).
CC -!- SIMILARITY: IDENTICAL FOR THE FIRST 132 AA, AND 75.3% IDENTICAL FOR THE NEXT 145 AA TO THE RNA1 POLYPROTEIN.
CC -!- SIMILARITY: TO THE RNA2 POLYPROTEIN OF OTHER NEPOVIRUSES.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-122 is the initiator.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
EMBL; D12477; BAA02043.1; -
PIR; JQ1093; GNVVTR.
DR InterPro; IPR005054; Nepo_coat.
DR InterPro; IPR005305; Nepo_coat_C.
DR InterPro; IPR005306; Nepo_coat_N.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF03391; Nepo_coat; 1.
DR Pfam; PF03688; Nepo_coat_C; 1.
DR Pfam; PF03689; Nepo_coat_N; 1.
KW Polyprotein; Coat protein; Repeat.
FT CHAIN 1321 1882 COAT PROTEIN (POTENTIAL).
FT DOMAIN 554 698 2.5 X TANDEM REPEATS, PRO-RICH.
FT REPEAT 554 606 1.
FT REPEAT 607 659 2.
FT REPEAT 660 698 3. (INCOMPLETE AND APPROXIMATE).
FT SEQUENCE 1882 AA; 206802 MW; 0F8958B63AE8DD9D CRC64;
Query Match 4.1%; Score 95.5; DB 1; Length 1882;
Best Local Similarity 20.3%; Pred. No. 27;
Matches 84; Conservative 50; Mismatches 129; Indels 151; Gaps 24;
QY 33 WYSTAPMATWAGGYDVHAQYKFLC---IHREVIIP-----ALGPVPE 74
DB 608 WSSPPLFANFKVNRG-----ACFLQVLPORVLDDECMDLLSLFQDLPEGPLPS 658
QY 75 KGQPMHMKSHLTRFGLPFELSFNYSK-----SLRFAFEPIGSLTG 115
DB 659 ---PSWSS-----PLPLFASFKNRGACFLQVLPAKVVSDFMVDLPFLFSP--VSH 707
QY 116 TKDDPFTQATRPVLQDL---KAMVPGLDLE-WDFHTKALVYSEEARLLDORDIEIP 170
DB 708 QEEPEMVPAYLEADSVDGITEAFFDLECEFSYDSYS-----DEEEAEW---ASVP 757
QY 171 VKTKONKLAADLEPSGDIVLKYIYPRKSIATGTPKRLMFDPAIKADKFGKVPATPLAI 230
DB 758 RCKTMSLCASITLADG-----AELGRKSHGVF-----LKRLVTVLQS 795
QY 231 LEE-FIASRAPTLGHFLSCDLVKP-----SESRIKVCWEROOLD---LASTEGITWL 279
DB 796 FEEPLYSSRA-----FYSVK-VKPVYRPFKFEHIDCTCLDGNNGEWESVDAMWRC 848
QY 280 NGR-----RNDPETLDGI-----DALRELWOLLAPVTE-----GLCPUN 313
DB 849 PGLRLTKRTFTTRDDWQVQLRIGFNGEGRYRNRVNLNLEEMDLSLHEYPEISSAPVQS 908

```

STRAIN=MOA; Serizawa K.;
Submitted (OCT-1999) to the EMBL/GenBank/DDJB databases.
-!- FUNCTION: IRBP shuttles 11-cis and all trans retinoids between the
retinal isomerase in the pigment epithelium and the visual
pigments in the photoreceptor cells of the retina.
-!- SUBCELLULAR LOCATION: Interphotoreceptor matrix that permeates the
space between the retina and the contiguous layer of pigment
epithelium cells.
-!- TISSUE SPECIFICITY: Expressed in the photosensitive tissues;
retina and pineal gland.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/ or
send an email to license@isb-sib.ch).
-----
EMBL; AF126968; AAA39331.2; -.
EMBL; AB033711; BAA85872.1; -.
MGD; MGI:97878; Rbp3.
InterPro; IPR003581; TSPC.
Pfam; PF02692; IRBP; 2.
SMART; SM00245; TSPC; 2.
Vitamin A; Transport; Repeat; Signal; Glycoprotein.
SIGNAL 1
CHAIN 17
DOMAIN 18 >610
REPEAT 18 >610
CARBOHYD 321 >610
N-LINKED (GLCNAC. . ) (POTENTIAL).
CARBOHYD 107 107
N-LINKED (GLCNAC. . ) (POTENTIAL).
CARBOHYD 205 205
N-LINKED (GLCNAC. . ) (POTENTIAL).
CONFLICT 513 513
G -> S (IN REF. 1).
NON TER 610 610

```

RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
other Chlamydia strains based on whole genome sequence analysis."
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-histidine + tRNA(His) = AMP +
diphosphate + L-histidyl-tRNA(His).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001648; AAD18802.1; -.
DR EMBL; AE002172; AAF3970.1; -.
DR EMBL; AB017159; AAF98618.1; -.
DR PIR; H72052; H72052.
DR HSSP; O32422; IQEO.
DR TIGR; CF0084; -.
DR HAMAP; MF_00127; -.
DR InterPro; IPR004154; HGTP anticodon.
DR InterPro; IPR004516; HisS.
DR InterPro; IPR002314; tRNA-synt 2b.
DR InterPro; IPR006195; tRNA ligase II.
DR Pfam; PF03129; HGTP-anticodon; 1.
DR Pfam; PF00587; tRNA-synt 2b; 1.
DR TIGRFAMs; TIGR00442; HisS; 1.
DR PROSITE; PS00862; AA tRNA LIGASE II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 430 AA; 49550 MW; 1892BB7E7C9B74 CRC64;

Query Match 4.1%; Score 94.5; DB 1; Length 430;
Best Local Similarity 22.7%; Pred. No. 4.2;
Matches 64; Conservative 41; Mismatches 126; Indels 51; Gaps 15;

QY 25 SNKDHKWMYSTAPMEATMAGAG-YDVHAYQKFLCI-----HREVIIPALGPYPEKG 76
DB 104 SHSDNKFY-IILMFYERQOAGRQHQFVEAIGVHPLRDAEVLALLWDFSVRG 162

QY 77 QPMHWKSHLFRGLFELSFNYSKSLRFAPEPLGSLTGKDPFNQAIR-----PVL 130
DB 163 L-QHMOIQNLFLG-GSETRFYDKVLRAVLEKSMGELSALSQQRFSTNVLRLDSKEPED 220

QY 131 QDLKAMVPGU-----DLEWPHFTKALVSVSEEARLLDRDIEIPVKTKQNKLAADLE 183
DB 221 QEIIRQAPPLIDYVSDLDKYNELDAL-----RVLEIP-VAINPLRVRGLD 267

QY 184 PSGDIVLK-TYIYPRIKSIATGPKERLMPDAIKAADKFGKVATPLAILEEFTIAERAPTL 242
DB 268 YSGLVFEATTTQEVSYALGGGR---YDGLISA--FGASLPACGFGVGLERAIQTL 321

QY 243 LGHFLSCDLVKPS-ESRIKVCYMERQDLASIEGIWTLNGRR 283
DB 322 LAQ-----KRIEPOPHKRLRIPMPDADQFCLG--WSOHLRR 357

RESULT 9
POL_HTL1A STANDARD; PRT; 896 AA.
AC P03362;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE POL polyprotein [Contains: Reverse transcriptase (EC 2.7.7.49);
DE Ribonuclease H (EC 3.1.26.4)].
GN POL.
OS Human T-cell leukemia virus type I (strain ATK) (HTLV-I).
OS Viruses; Retroviridae; Retroviridae; Deltaretrovirus.

OX NCBI_TaxID=11926;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=83221647; PubMed=6304725;
RA Seiki M., Hattori S., Hirayama Y., Yoshida M.;
RT "Human adult T-cell leukemia virus: complete nucleotide sequence of
the provirus genome integrated in leukemia cell DNA."
RL Proc. Natl. Acad. Sci. U.S.A. 80:3618-3622(1983).
CC
CC SEQUENCE OF 69-185 FROM N.A.
RX MEDLINE=89210803; PubMed=2468487;
RA Bangham C.R.M., Daenke S., Phillips R.E., Cruickshank J.K.,
RA Bell J.I.;
RT "Enzymatic amplification of exogenous and endogenous retroviral
sequences from DNA of patients with tropical spastic paraparesis."
RL EMBO J. 7:4179-4184(1988).
CC
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
phosphomononucleotide.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ {DNA} (N).
CC -!- PTM: Cleavage sites that yield the mature proteins remain to be
determined.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; J02029; AAA96673.1; -.
DR EMBL; X14144; CAA32360.1; -.
DR PIR; A03961; GNLJGH.
DR HSSP; P03355; 1MML.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; Rvise.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF00075; rna5H; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
DR Polyprotein; Hydrolase; Endonuclease; Nuclease; Transferase;
KW RNA-directed DNA polymerase.
SQ SEQUENCE 896 AA; 100141 MW; 113D45D4BD79C65F CRC64;

Query Match 4.1%; Score 94.5; DB 1; Length 896;
Best Local Similarity 20.7%; Pred. No. 12;
Matches 79; Conservative 62; Mismatches 151; Indels 89; Gaps 18;

QY 71 PYPEKGQPMHWKSHLFRGLFELSFNYSKSLRFAPEPLGSLTGKDDP--ENTQAIRP 128
DB 123 FLPKQFQY-----FAFTVPOOCNYGPG-TRYAMKVLFPQ--GFRNSPTLFEMQ-LAH 170

QY 129 VLQDLKAMVPGU-LWEPDHFHTKA-----LVVSEEARLLDRDIEIPVKTKQNKLA 180
DB 171 ILQPIRQAPFQCTILQYVDDILLASPSHEDLLILLSEATWASLHGLFVSENKTQTPCT 230

QY 181 DLEPSGDIVLTYIYPRIKSIATGPKERLMPDAIKAADKFGKVATP--LAILEE--PIA 236
DB 231 -----IKFLQIISPNHLYDAVPTPIRSRWALPELQALLGFIQWVS 273

QY 237 ERAPTLIG--HFLSCDLVKPSERIKVCYMERQL--DLASIEGIWTLNGRRNDPDTLGLD 293
DB 274 KGTPTLRQLRPLSLYCALORHDPDRDIYLNPSQVSLVQLRQALSONCERSLVQLTLLG 333

QY 294 ALREWLQPLPVTEGLCPLENCFYEPGTS-----PQEQLPFI-INFTLSPKALPEPQIYF 347
DB 334 AI-----MLTLT-----GTTTVVFQSQEQWPLVWLHAPLEPTSQCPWQGLIA 375

QY 348 PAFGQNDKTIAGLATFFESRCGWGLAKSYPADLASVYPDVDTQATNHLQAWISFYK 407

```
Db 376 SAVLLDK-----YTLQSYGLLQTIHNIHQFQTSHPV----- 417
QY 408 KPYMSVYLHTPEAFSAQAQEV 428
Db 418 -PILHSHRPFKNGAQTGEL 437

RESULT 10
SVY_SVY3
ID SVY SVY3 STANDARD; PRT; 910 AA.
AC Q55522;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).
GN VALS OR SLR0557.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiura M., Tabata S.
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
CC + L-valyl-tRNA(Val).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; D64006; BAA10881.1; -
DR PIR; S76034; S76034.
DR HSP; P96142; IGAX.
DR InterPro; IPR002300; tRNA-synt_la.
DR InterPro; IPR001412; tRNA-synt_I.
DR InterPro; IPR002303; tRNA-synt_val.
DR InterPro; IPR009008; ValRS_lleRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TENASYNTHVAL.
DR TIGR; TIGR00422; vals; 1.
DR PROSITE; PS00178; AA_tRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
Complete proteome.
FT SITE 46 56 "HIGH" REGION.
FT SITE 539 543 "KMSKS" REGION.
FT BINDING 542 542 ATP (By similarity).
FT BINDING 542 542 ATP (By similarity).
SQ SEQUENCE 910 AA; 102737 MW; 4D909BCD889BF42E CRC64;

Query Match 4.0%; Score 94; DB 1; Length 910;
Best Local Similarity 22.1%; Pred. No. 13;
Matches 83; Conservative 46; Mismatches 117; Indels 130; Gaps 22;

QY 63 EVIIPALGPYKPGQPMHKSHTL---RGLPEFLSFNYSKSLRFAPEPLSLTGTGDD 119
Db 262 ELVDPEFGGCVKVPFAHPDNPVFGQRHNLFPINLNDGSL-----NENGDFAG--QD 315
QY 120 PFNTQAIRVLODKAMVPGDLDFHFTKALVYSEEEARTLL---DRDIEIPVFKTON 176
Db 316 RF--EARKNVQALEA-----QGLFKVISPYRHSVPYDGRG-KVPV----- 353
QY 177 KLAADLEPSGDIVLXTYIYPRIKSIATGTPKRLMFDALKAADKPKVATPLALEZ--- 233
```

```
Db 354 -----EP-----LISQWFKIESLAQNA-----LACDEDSN 381
QY 234 --FIAERAPTLIGHFLSCDLVKPSESRIKVCYMERQDLASIEGIWTNGRNDPFDLDG 291
Db 382 PNFVPERGWKVYRDWL-----VKLKWICISQJL-----WM--GHQ----- 414
QY 292 LDALRELWQLLPVTEGJLCPNCFYEPGTSQEQOLPFIINFTLSPKALPFPQIYFFAPG 351
Db 415 ----IPAWYIVSETNG-----AITDHTPFIVAYDEAEALAKAKAE-YGPTVQ 456
QY 352 -QNDKTIAGELATPFESRGWGLAKSYPA---DLASYVPDVLQTA-NHQAWI----- 400
Db 457 LQDDPDV---LDTWFSGLWFFSTMGWPEQTDLLAKYYPSTLVGTGDIFFWVARMVMM 513
QY 401 SFSYKGGKPYMSVYLH 416
Db 514 AGHFTGQIPFXDVYIH 529

RESULT 11
SVL_STRPN
ID SVL_STRPN STANDARD; PRT; 833 AA.
AC Q97SS0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR SP0234.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetelin H., Nelson K.B., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen, C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:1498-506(2001).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AE007338; AAK74433.1; -
DR PIR; H95029; H95029.
DR TIGR; SP0254; -.
DR HAMAP; MF_00049; -.
DR InterPro; IPR002302; Leu-tRNA-syntla.
DR InterPro; IPR002300; tRNA-synt_la.
DR InterPro; IPR001412; tRNA-synt_I.
DR InterPro; IPR009008; ValRS_lleRS_edit.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TENASYNTHLEU.
DR TIGR; TIGR00396; leus_bact; 1.
DR PROSITE; PS00178; AA_tRNA_LIGASE_I; 1.
```

KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome. 52 "HIGH" REGION.
FT SITE 610 614 "KMSKS" REGION.
FT SITE 613 613 ATP (BY SIMILARITY).
FT BINDING 613 613
SQ SEQUENCE 833 AA; 94421 MW; 7CAA4FB5B4F99E83 CRC64;
Query Match 4.0%; Score 93; DB 1; Length 833;
Best Local Similarity 19.1%; Pred. No. 14;
Matches 95; Conservative 56; Mismatches 188; Indels 148; Gaps 26;
QY 16 YVLSQALN-SNKHDTKW--WYSTAMPATMAGAGYDVHQAQYKFL-----CHREVIIP- 67
Db 119 YDMDREVTITDNYKWTQWIFT-----KLVEKGLAYEAEPVNVNVEELGTALANEVLPD 174
QY 68 ---ALGPYKPGQM-HWKSHTLRFPG--LPPEL-SFNYSKSLRFAFEPILGSLTGTK--- 117
Db 175 GTSEGGYVVRKPMQKMLKITAYAERLNDLDELWSESIDMQRNWIGKSTGANVTF 234
QY 118 ----DDPENTQAIRPVL---QDLKAMVPGDLDFHFTKALVVSEEEARTLLDRDIEI 169
Db 235 KVKGTCKEFTVTRPTDPLFGATFTVLAP-----EHELVDAITSEQAQAVAD----- 282
QY 170 PVFQTKNLAAIDLEPSGDIVLKYIYPRKSIATGTPKERMFPDAIKAADKKGKATPLA 229
Db 283 --YHQASLKSRLDRLAKETGVTGAYAINPVNKGEMPIWIADYVLASYGTGAV--- 337
QY 230 ILBEFTAERAAPTLLGHFLSCDLVKPSESRIKVCMERQLDLASI-----EGITWLN 280
Db 338 -----MAVPAHQ--RDWEFAKQDPLPVEVLEGGNVEEAAYTED 375
QY 281 GRNDPPTLGL--DALREL--WQLPVTEGLC-----PLPN 313
Db 376 GLHVNDFDLGLNKDAIAKIVAW---LEEKGGQGVKTYRLRDLWFSRQRYWGEPIPI 431
QY 314 CFVEPGTS---PQEQLPFIINFLSPKSALEPQIYFPAQGND---KTAIEGLATFPFS 367
Db 432 IHMEDGTSVAPTELEPLVPVT---KDIRPSTGSGPLANLTDWLEVTREDGVKGRRT 488
QY 368 R---GWGGLAKSY-----PADLASYPVDVLDQTAN-----HL---QAWISF 402
Db 489 NTPEQWAGSSWYLYRVIDPHNTEKLADEDLKQWLPVDIVGGAHVLHLLVAREWHKF 548
QY 403 SY-----KGKPYMSVY 414
Db 549 LYDLGVVPTKEPFQKLF 565

RESULT 12
POLN_EBVT STANDARD; PRT; 2492 AA.
AC P27282;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural polyprotein [Contains: Nonstructural protein NSP1;
DE Nonstructural protein NSP2; Nonstructural protein NSP3; Nonstructural
DE protein NSP4].
OS Venezuelan equine encephalitis virus (strain Trinidad donkey).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
CX NCBI_TaxID=11038;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89243175; PubMed=2524126;
RA Kinney R.M., Johnson B.J.B., Welch J.B., Tsuchiya K.R., Trent D.W.;
RT "The full-length nucleotide sequences of the virulent Trinidad donkey
RT strain of Venezuelan equine encephalitis virus and its attenuated
RT vaccine derivative, strain TC-83,"
RL Virology 170:19-30 (1989).
CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS

CC BETWEEN THE CODONS FOR 1879-GLN AND 1880-ARG.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; J04332; AB02518.1; --
DR PIR; A31467; MNWVTD.
DR MEROPS; C09.001; --
DR InterPro; IPR002589; Alpp.
DR InterPro; IPR002620; Peptidase C9.
DR InterPro; IPR001788; RNA dep rNAPol2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PS_vir.
DR InterPro; IPR000606; Viral_helicase.
DR Pfam; PF01661; Alpp; 1.
DR Pfam; PF01707; Peptidase C9; 1.
DR Pfam; PF00978; RNA dep rNAPol2; 1.
DR Pfam; PF01443; Viral_helicase; 1.
DR SMART; SM00506; Alpp; 1.
KW Polyprotein; Nonstructural protein; RNA-binding; ATP-binding;
KW Helicase.
FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 1330 1879 NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 1880 2492 NONSTRUCTURAL PROTEIN NSP4.
FT NP_BIND 721 728 ATP (POTENTIAL).
SQ SEQUENCE 2492 AA; 277902 MW; 1BAD415B70DC3FA0 CRC64;
Query Match 4.0%; Score 93; DB 1; Length 2492;
Best Local Similarity 20.9%; Pred. No. 64;
Matches 67; Conservative 45; Mismatches 91; Indels 118; Gaps 18;
QY 57 FLC-IHREVIIPALGPYEPKQPMHKSHTLRFGLP-----FELSEFNYSKSLRFAFEP- 109
Db 2206 YLCGHRELIV-----RRLNAVLLPNHITLFDMSAEDFADIAEHFPG 2248
QY 110 -----LGSUTGTDKDPFNTQAIRPVLDKAMVPGDLDFHFTKALVVSEEEARTL 162
Db 2249 DCVLETDIASFDKSEDDAMALTALM-ILEDL-----GVDAELL-----TL 2287
QY 163 LD-----RDIEIPVFKTONKLAADLEPSG-----DIVLKYIYPR- 198
Db 2288 IEAFAFGESSIHLPT-KTKFKGAMMK-SGMFLTLFVNTVINIVIASVLRERLTGSPCA 2345
QY 199 -----KSIATGTPKERLMFDAIKADKFGKQVAT-----PLAILEBFIABERAPTLGHFLSC 249
Db 2346 AFIGDDNIVKGVSKDLNAD-----RCATWLNMEVKIIDA VVGKAPVFCGFI LC 2396
QY 250 DLVKPSEERI-----KYCYMERQLDLASIEGIWTLNGRNDPFDLGLDALRE---LWQL 301
Db 2397 DSVTGTACRVADPLKRLFKLGLKPL-----AADDEHDDRRRALHEESTRWNR 2443
QY 302 LPVTEGLCLPNCFYEP-GTS 321
Db 2444 VGILSELKAVESRYETVGTG 2464
RESULT 13
ADHE_CLOAB STANDARD; PRT; 862 AA.
AC P33744; Q45808; Q45809;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde-alcohol dehydrogenase [includes: Alcohol dehydrogenase
DE (EC 1.1.1.1) (ADH); Acetaldehyde dehydrogenase [acetylating]
DE (EC 1.2.1.10) (ACDH)].
GN ADHE OR AAD OR CAP0162.

DR InterPro: IPR006708; Pex19.
DR Pfam: PF04614; Pex19; 1.
KW Hypothetical protein; Peroxisome; Prenylation; Lipoprotein.
FT DOMAIN 270 275 Poly-ALA.
FT LIPID 279 279 S-farnesyl cysteine (Potential).
SQ SEQUENCE 282 AA; 30988 MW; 681FDF8F08DBE4 CRC64;

Query Match 4.0%; Score 92; DB 1; Length 282;
Best Local Similarity 23.1%; Pred. No. 3.7; Mismatches 104; Gaps 16;
Matches 67; Conservative 37;

QY 154 VSEEARLTLDRIEIPVKTQNKLAADLEPG-GDIVLKYIYPIKSIAGTTPKERLMP 212
DB 1 MTDETTQNIKD-----KTEELAALDQTLGEFT-----ATPAPKPRTTD 40

QY 213 D-----ATKADKKGKQVATPLAILEEPIAERAPLLGHFLSCDLVKPSESRI 259
DB 41 DELDELMASDAQEAQAKADQKMLEQNVTLQEEAMKAG-----ADPSEGE- 88

QY 260 KVCYMERQLDASIEGIWTLNGRRNDPDTLDGLDALRELWQL-----PVTEGLC 309
DB 89 -----GQQLD-----PNDPEALAMMDALKQLMCESSNVANASNEEFMAGLD 131

QY 310 PL--PNCFYEPGTSPOEQLPFFINFTLSPKSL--PEPOIY--FPAP-----GQNDKT- 356
DB 132 MLRSPN-----SPMEPTFMIMQTLASKEVYVPLKEIFDNYPKYLEDNAGLDATK 184

QY 357 -----IAEGLATPFESRGMGLAKSYPADLAS-YPDVDLQTAHLQ 397
DB 185 ERYEKQFEVLGKCTEFEXQ--PELAEVQPVDAATQAPADPASTIEHFE 232

DR HSP; P32168; IE94.
DR HAMAP; MF 00249; -; 1.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003593; AAA_ATPase.cent.
DR InterPro: IPR001270; Chaperonin c1pa/B.
DR InterPro: IPR004491; Hsp_HslVU.
DR Pfam; PF00004; AAA; 1.
DR PRINTS; PR00300; CLPPTROTEASEA.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00390; hslU; 1.
KW Chaperone; ATP-binding; Complete proteome.
FT NP_BIND 59 66 ATP (POTENTIAL).
SQ SEQUENCE 443 AA; 50164 MW; 6AF0F90849B686DA CRC64;

Query Match 4.0%; Score 92; DB 1; Length 443;
Best Local Similarity 24.6%; Pred. No. 7;
Matches 73; Conservative 44; Mismatches 116; Indels 64; Gaps 17;

QY 119 DPFTQAIRPVLDLKAMPV-GLDLEWFDHFTKALVSEEE-ARTLLD-RDIRIPVFKTQ 175
DB 123 DKIEEAVVEKIAKLLPFLPSGVSEKKQBYANSLKQOQRIAGGELDSREIEIEVRKS 182

QY 176 NKLAADLEPSGDIVLKTYIYPIKSIATGTPKERLMFDAIKAADKFGKVATPLAILEEFT 235
DB 183 IEIDSNVFPB---ILRV-----QENLIKVFHKEQDKVKKTLKSVKEAKEALK 225

QY 236 AERAPTL-GHFLSCDLVKPSESRIKYCMERQLDASIEGIWTLNGRRNDPDTLDGLDA 294
DB 226 AEISDTLLDGEAIKVEGLKRAESSGVIFIDE--IDKIAVS---SKEGSRQDP-SKEGVQR 279

QY 295 LRELQQLPVTEGLCPNCFYEPGTSPOEQLPFI--NFTLS-PKSALPEPQIYFP--- 348
DB 280 -----DLLPIVEG--SVNTKY--GSIKTEHILFIAAGAFHLSKPSDLPELQGRFPLRV 330

QY 349 -----AFQNDKTI-----AEGLATPFESRGMGLAK-SYPADLAS 383
DB 331 ELENLTBEIMVMILTQTKTSIIKQYQALLKVEGVEIAFEDDAIKELAKLSYNANQKS 387

Search completed: July 22, 2004, 18:45:16
Job time : 22 secs

DR InterPro: IPR006708; Pex19.
DR Pfam: PF04614; Pex19; 1.
KW Hypothetical protein; Peroxisome; Prenylation; Lipoprotein.
FT DOMAIN 270 275 Poly-ALA.
FT LIPID 279 279 S-farnesyl cysteine (Potential).
SQ SEQUENCE 282 AA; 30988 MW; 681FDF8F08DBE4 CRC64;

Query Match 4.0%; Score 92; DB 1; Length 282;
Best Local Similarity 23.1%; Pred. No. 3.7; Mismatches 104; Gaps 16;
Matches 67; Conservative 37;

QY 154 VSEEARLTLDRIEIPVKTQNKLAADLEPG-GDIVLKYIYPIKSIAGTTPKERLMP 212
DB 1 MTDETTQNIKD-----KTEELAALDQTLGEFT-----ATPAPKPRTTD 40

QY 213 D-----ATKADKKGKQVATPLAILEEPIAERAPLLGHFLSCDLVKPSESRI 259
DB 41 DELDELMASDAQEAQAKADQKMLEQNVTLQEEAMKAG-----ADPSEGE- 88

QY 260 KVCYMERQLDASIEGIWTLNGRRNDPDTLDGLDALRELWQL-----PVTEGLC 309
DB 89 -----GQQLD-----PNDPEALAMMDALKQLMCESSNVANASNEEFMAGLD 131

QY 310 PL--PNCFYEPGTSPOEQLPFFINFTLSPKSL--PEPOIY--FPAP-----GQNDKT- 356
DB 132 MLRSPN-----SPMEPTFMIMQTLASKEVYVPLKEIFDNYPKYLEDNAGLDATK 184

QY 357 -----IAEGLATPFESRGMGLAKSYPADLAS-YPDVDLQTAHLQ 397
DB 185 ERYEKQFEVLGKCTEFEXQ--PELAEVQPVDAATQAPADPASTIEHFE 232

DR HSP; P32168; IE94.
DR HAMAP; MF 00249; -; 1.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003593; AAA_ATPase.cent.
DR InterPro: IPR001270; Chaperonin c1pa/B.
DR InterPro: IPR004491; Hsp_HslVU.
DR Pfam; PF00004; AAA; 1.
DR PRINTS; PR00300; CLPPTROTEASEA.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00390; hslU; 1.
KW Chaperone; ATP-binding; Complete proteome.
FT NP_BIND 59 66 ATP (POTENTIAL).
SQ SEQUENCE 443 AA; 50164 MW; 6AF0F90849B686DA CRC64;

Query Match 4.0%; Score 92; DB 1; Length 443;
Best Local Similarity 24.6%; Pred. No. 7;
Matches 73; Conservative 44; Mismatches 116; Indels 64; Gaps 17;

QY 119 DPFTQAIRPVLDLKAMPV-GLDLEWFDHFTKALVSEEE-ARTLLD-RDIRIPVFKTQ 175
DB 123 DKIEEAVVEKIAKLLPFLPSGVSEKKQBYANSLKQOQRIAGGELDSREIEIEVRKS 182

QY 176 NKLAADLEPSGDIVLKTYIYPIKSIATGTPKERLMFDAIKAADKFGKVATPLAILEEFT 235
DB 183 IEIDSNVFPB---ILRV-----QENLIKVFHKEQDKVKKTLKSVKEAKEALK 225

QY 236 AERAPTL-GHFLSCDLVKPSESRIKYCMERQLDASIEGIWTLNGRRNDPDTLDGLDA 294
DB 226 AEISDTLLDGEAIKVEGLKRAESSGVIFIDE--IDKIAVS---SKEGSRQDP-SKEGVQR 279

QY 295 LRELQQLPVTEGLCPNCFYEPGTSPOEQLPFI--NFTLS-PKSALPEPQIYFP--- 348
DB 280 -----DLLPIVEG--SVNTKY--GSIKTEHILFIAAGAFHLSKPSDLPELQGRFPLRV 330

QY 349 -----AFQNDKTI-----AEGLATPFESRGMGLAK-SYPADLAS 383
DB 331 ELENLTBEIMVMILTQTKTSIIKQYQALLKVEGVEIAFEDDAIKELAKLSYNANQKS 387

Search completed: July 22, 2004, 18:45:16
Job time : 22 secs

RESULT 15

HSLU HELPJ STANDARD; PRT; 443 AA.
ID HSLU HELPJ
AC Q9ZLW1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent hsl protease ATP-binding subunit hslU.
GN HSLU OR JHP0465.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=99120557; PubMed=9923682;
RX Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori";
RL Nature 397:176-180(1999).
CC -!- FUNCTION: Chaperone subunit of a proteasome-like degradation
complex (By similarity).
CC -!- SUBUNIT: A double ring-shaped homohexameric of hslU is capped on
each side by a ring-shaped hslU homohexameric (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the clpx chaperone family. HslU subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001480; AAD06034.1; -.
DR PIR; F71929; F71929.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:41:02 ; Search time 39 Seconds
(without alignments)
3535.422 Million cell updates/sec

Title: US-10-099-704-2
Perfect score: 2322
Sequence: 1 MEISKKAATLIPKPYVLSQ.....PEAFSAQAQEVANCHDGNP 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriopl.*
17: sp_archaeap.*

ALIGNMENTS

RESULT 1

O94204 PRELIMINARY; PRT; 448 AA.
AC O94204;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE Dimethyl-allyl-tryptophan-synthase.
OS Claviceps purpurea (Ergot fungus).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae; Clavicipes.
OX NCBI_TaxID=5111;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P1;
RX MEDLINE=99168777; PubMed=10071219;
RA Tudzynski P., Hoelter K., Correia T.H., Arntz C., Grammel N.,
RA Keller U.;
RT "Evidence for an ergot alkaloid gene cluster in Claviceps purpurea.";
RL Mol. Gen. Genet. 261:133-141(1999).
DR EMBL; AJ011963; CAB39314.1; .
SQ SEQUENCE 448 AA; 51672 MW; 625709FAAF269242 CRC64;

Query Match 39.3%; Score 911.5; DB 3; Length 448;
Best Local Similarity 43.7%; Pred. No. 9.5e-69;
Matches 181; Conservative 71; Mismatches 151; Indels 11; Gaps 4;
QY 15 FYVLSQLNLGNKDHTRKWTSTAPMFATMAGAGYDVAHQYKFLGCIHREVIIPALGPYPE 74
Db 12 YEILSLIFDFPSNEQRLWHSTAPMFAAALDNAGYNIHQVRLHGFKKHIIIFLGVIPT 71
QY 75 KQCPMHVKSILTRFGLPFELSFNYSKSLIRFAFPLGSLTGTGKDPFNTQAIRPVLQDLK 134
Db 72 KDKK-RWLSILTRGCLPLELSLNCDSVVRVTEPINEVTGTEKDTFTNLAIMTSVQKLA 130
QY 135 AMVPGLDLEWDFHFKALVWSEEEARTLLDRDIEIPFVKTONKLAADLEPSGDIVLKTYI 194
Db 131 QIQAGIDLEWFSYFKDELTLDESEATLQSNELVKEQIKTONKLAADLEKES-QPALKVVF 189

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	911.5	39.3	448	3	O94204 claviceps p
2	908.5	39.1	448	3	Q9C141 claviceps p
3	895.5	38.6	455	3	Q12594 claviceps f
4	849.5	36.6	441	3	Q9C140 claviceps p
5	202.5	8.7	187	3	Q9C451 penicillium
6	117	5.0	577	10	Q8S922 oryza sativ
7	112	4.8	1772	3	Q96X17 pichia past
8	110.5	4.8	462	17	Q8TXF1 methanopyru
9	107.5	4.6	490	5	Q21549 caenorhabdi
10	104	4.5	595	5	Q7YZB8 tenebrio mo
11	102	4.4	1043	4	Q86W25 homo sapien
12	101	4.3	503	11	Q8KOC4 mus musculu
13	101	4.3	503	11	Q8BSQ7 mus musculu
14	101	4.3	878	10	Q42710 cucumis sat
15	101	4.3	1621	16	Q9KTA5 vibrio chol
16	100	4.3	323	16	Q9A9D7 caulobacter

```
QY 195 YPRIKSIATGTPKRLMFDAL-KAADKFGKVATPLAILEBFIARAPT-----LLGH 245
Db 190 YPHLKSIATGTHDLIFDSVLSQSHDSIQAFQVLCDYVSRRHSAEVDQHGALHAR 249
QY 246 FLSCDLVKPESRIKVCYMERQOLDASIEGIWTLNGRNDPETHLGLDALRELWQLLPVT 305
Db 250 LLSCDLIDPAKSRVKIYLKXTVLSWMDLWLGQGVDAWMDGLDMLRELWSLLKVP 309
QY 306 EGLCPLNCFYEPGTSQBOLPFIINFTLSPKALPEPQIYFPAFQNDKNTIAEGLATFF 365
Db 310 TGHLEYPKGYLEGEIPNEQLPSMANYTLHNDPMPPEQVYFVFGNDAEISNALTIF 369
QY 366 ESRGWGLAKSYPADLASYPVDLOTANHLQAWISFSYKGGKPYMSVYLHTE 419
Db 370 QRHGFDDMAKNYRVFLQDSYPYHDFESLNYLHAYISFSYRNKPYLSVYLHTE 423

RESULT 2
Q9C141
ID Q9C141 PRELIMINARY; PRT; 448 AA.
AC Q9C141
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DE Putative dimethyl-allyl-tryptophan-synthase.
GN CPD2.
OS Claviceps purpurea (Ergot fungus).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
OX NCBI_TaxID=5111;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T5;
RA Antz C., Tudzynski P.;
RT "Molecular analysis of dimethyl-allyl-tryptophan-synthase-genes.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBDatabases.
DR EMBL; AJ312753; CAC37396.1; -.
SQ SEQUENCE 448 AA; 51520 MW; 1ED9F8900810B59 CRC64;

Query Match 39.1%; Score 908.5; DB 3; Length 448;
Best Local Similarity 44.0%; Pred. No. 1.7e-68;
Matches 182; Conservative 65; Mismatches 156; Indels 11; Gaps 4;

QY 15 FVYLSQALNLSNKHDKTWYSTAPMTWAGAGYDVHQAQKFLCIHREVIIPALGPYPE 74
Db 12 YEILSLIFDPPSNEQRLWHSAPMAALDNAGYSVHDQYRHLISFKTHIIPFLGVYPT 71
QY 75 KGOPMHWKSHLTFGLPPELSFNYSKSLRFAPEPLGSLTGTGKDPFNTOAIRPVLDLK 134
Db 72 KGQE-RWLSILTRCGFLPSELSNCTDSVRYAYEPINEMTGKDPSTNLTPIIGSVQKLA 130
QY 135 AMVPGDLLEWDFHFTKALVSEBEARTLLDRDIEIPVFTQNKLAADLEPSGDIVLITYI 194
Db 131 QIQAGIDLEWFSYFKDELTDSESAILODELVEKQIKTQNKALDLKES-QFALKYVF 189
QY 195 YPRIKSIATGTPKRLMFDAL-KAADKFGKVATPLAILEBFIARAPT-----LLGH 245
Db 190 YPHLKSIATGTHDLIFDSVLSQSHDSIQAFQALCDYVSRRNDSSEVDQHGALHAR 249
QY 246 FLSCDLVKPESRIKVCYMERQOLDASIEGIWTLNGRNDPETHLGLDALRELWQLLPVT 305
Db 250 LLSCDLIDPAKSRVKIYLKXTVLSWMDLWLGQGVDAWMDGLDMLRELWSLLKVP 309
QY 306 EGLCPLNCFYEPGTSQBOLPFIINFTLSPKALPEPQIYFPAFQNDKNTIAEGLATFF 365
Db 310 TGHLEYPKGYLEGEIPNEQLPSMANYTLHNDPMPPEQVYFVFGNDAEISNALTIF 369
QY 366 ESRGWGLAKSYPADLASYPVDLOTANHLQAWISFSYKGGKPYMSVYLHTE 419
Db 370 QRHGFADMAKNYRVFLQDSYPYHDFESLNYLHSLVFSYRNKPYLSVYLHTE 423
```

```
RESULT 3
Q12594
ID Q12594 PRELIMINARY; PRT; 455 AA.
AC Q12594
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Dimethylallyl-trans-transferase (EC 2.5.1.1).
GN DMAW.
OS Claviceps fusiformis (Ergot fungus).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
OX NCBI_TaxID=40602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96067540; PubMed=7468077;
RA Tsai H.F., Wang H., Gebler J.C., Poulter C.D., Schardl C.L.;
RT "The Claviceps purpurea gene encoding dimethylallyltryptophan
RT synthase, the committed step for ergot alkaloid biosynthesis.";
RL Biochem. Biophys. Res. Commun. 216:119-125(1995).
DR EMBL; L39440; AAC18893.1; -.
DR GO; GO:0004161; F:dimethylallyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
SQ SEQUENCE 455 AA; 51858 MW; A913F8B60EBBEB63 CRC64;

Query Match 38.6%; Score 895.5; DB 3; Length 455;
Best Local Similarity 42.1%; Pred. No. 2.2e-67;
Matches 187; Conservative 75; Mismatches 151; Indels 31; Gaps 9;

QY 3 ISKXAATLPPKPYVLSQALNLSNKHDKTWYSTAPMTWAGAGYDVHQAQKFLCIHR 62
Db 2 MTXAPATAV---YDTLSLLDFNFQEQRLWHSIAPMAALDTAGNVHDQYEHGIFK 58
QY 63 EVIIPALGPPEKQOPMH-KWSHLTRFGLPPELSFNYSKSLRFAPEPLGSLTGTGKDDPF 121
Db 59 KHIIFPLGVYPAQGG--HTWPSVLTRYGIPPELSNCLDSVVRVYTFEPTTEHTGTGDSY 116
QY 122 NTOAIRPVLDLKAMVPGLDLEWDFHFTKALVSEBEARTLLDRD-IEIPVFTQNKLA 180
Db 117 NAFALLECIQKLVRIQPGIDMENFSYFRELVLNATESARLGRNDSVNOQPIPTQNKAL 176
QY 181 DLEPSGD-IVKYYIYPIRIKSIATGTPKRLMFDAL-KAADKFGKVATPLAILEBFIAR 238
Db 177 DLK--GDRFALKVYLPHLKSIATGVSSHDLIENSVRKLSQKHTSIQPSFNVLCDYVASR 234
QY 239 -----AFTLGLHFLSCDLVKPESRIKVCYMERQOLDASIEGIWTLNGRR 283
Db 235 NDPDSNAEAEAGVPASALRARLLSCDLVDPSKRIKYLLEQVSLTAWEDLWTLGGR 294
QY 284 NDPETDGLDALRELWQLLPVTEGLCPNCFYEPGTSQBOLPFIINFTLSPKALPEP 343
Db 295 TDSSTLNGLDMMRELWHLQIPSGFMKYPESDLKGEVDEQLPSMVHYALHPDQMPPEP 354
QY 344 QIYFPAFGQNDKNTIAEGLATFFESRCGGGLAKSYPADLASYPVDLOTANHLQAWISFS 403
Db 355 QVTFYTFVFGMSDAGITNALATFFSRHGWEMAKYRVFLGSGFPHDFESLNYLHLYTYSFS 414
QY 404 YGKKPYMSVYLHTE-----AFS 422
Db 415 YRKNKPYLSVYLHSPETGQWPAFS 438

RESULT 4
Q9C140
ID Q9C140 PRELIMINARY; PRT; 441 AA.
AC Q9C140
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE Putative dimethyl-allyl-tryptophan-synthase (fragment).
GN CPD1.
```

OC	Eukaryota; Fungi; Ascomycota; Peizomycotina; Sordariomycetes;
OC	Hypocreomycetidae; Hypocreales; Clavicipitaceae; Clavicipes.
OX	NCBI_TaxID=5111;
RN	SEQUENCE FROM N.A.
RP	SEQUENCE FROM N.A.
RC	STRAIN=75;
RA	Correia T., Tudzynski P.;
RT	"Molecular analysis of dimethyl-allyl-tryptophan-synthase-genes.";
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AJ312754; CAC37397.1; --
FT	NON TER 441
SQ	SEQUENCE 441 AA; 50690 MW; ECD342B080D42FE CRC64;
Query Match 36.6%; Score 849.5; DB 3; Length 441;	
Best Local Similarity 41.8%; Pred. No. 1.7e-63;	
Matches 172; Conservative 70; Mismatches 158; Indels 11; Gaps 4;	
QY	15 EYVLSQALNLSKDHKXWYSTAPNFATMMAGAGYDVHQAQYKFLCIHREVIIPALGPYPE 74
DB	12 YEILSLIFDPNEQRLWHSTAPFAMLDNAGYNIHDOYHGLGIFKXHIIPFLGVYPT 71
QY	75 KGQPMHWSHLTRFGLPFELSNYSKSLRPAFEPGLSLTGKODPFNTQAIRPVLDLK 134
DB	72 KDKE-RWLSILTRGLPLSLNCTDSVVRVYTYEINVTGTEKDPFNTLAIMASVQKLA 130
QY	135 AMVPGLDLEWDFHTKALVUSSEEARLLDRDIEIPVFTONKLAADLEPSGDIVLKTVI 194
DB	131 QIQAGIDLEWFSYFKDELTDSESATLSQSNELVKEQIKTKNKLALDKES-QFALKVYF 189
QY	195 YPRIKSIATGTPKRLMFDAI-KAADKFGKVATPLAILEEFIAERAPT-----LIGH 245
DB	190 YPHLKSIAIGKSTHDLIFDSVFKLSQKSDSIQAPXQVLCVYVSRNNSAESQDIALHAR 249
QY	246 FLSCDLVKSEBIKIVYCHERQDLASIEGIWTLNGRRNDPDTLGLDALRELWQLLPVT 305
DB	250 LLESCDLIDPAKSVXYLLEKTVLSVMEDLWTLGGQVDASVMDGLDRLRLSLLKVP 309
QY	306 EGLCPLPNCPEYEGTSPQEQLPPIINFTLSPKSALPEPOLYFPAPFGQNDKTIAGELATPF 365
DB	310 TGHLFEPKGLVLEGEIPEQLPSMANYTLHNNPMPPEQVYTVFGMNDAEISNALTIFP 369
QY	366 ESRGWGLAKSPADIASVYDVLQTAHNLQAWISFYSYKGGKPKYMSVLH 416
DB	370 QREHGFDMAKKRVFLQDSYPYHDFESLNYLHAYISFYSYWPVGESTELYQH 420
RESULT 5	
ID	Q9C451 PRELIMINARY; PRT; 187 AA.
AC	Q9C451;
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	Dimethylallyl tryptophan synthase (Fragment).
GN	PAXD.
OS	Penicillium paxilli.
OC	Eukaryota; Fungi; Ascomycota; Peizomycotina; Eurotiomycetes;
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX	NCBI_TaxID=70109;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21106005; PubMed=11169115;
RA	Young C., McMillan L., Teifer E., Scott B.;
RT	"Molecular cloning and genetic analysis of an indole-diterpene gene
RT	cluster from Penicillium paxilli.";
EL	Mol. Microbiol. 39:754-764 (2001).
DR	EMBL; AF279808; AAK11526.1; --
FT	NON TER 1
SQ	SEQUENCE 187 AA; 21469 MW; 6D303E02B8977FA3 CRC64;
Query Match 8.7%; Score 202.5; DB 3; Length 187;	
Best Local Similarity 30.5%; Pred. No. 4.2e-09;	
Matches 50; Conservative 33; Mismatches 66; Indels 15; Gaps 5;	
QY	257 SRIKYCMERQDLASIEGIWTLNGRRNDPDTLGLDALRELWQLLPVTGLCPPLNCFY 316
DB	12 TRVKIYFATOSTAFNNVDIFILGRLDGFEMORATKELKLMW-----STVAIFDGLR 65
QY	317 EPGTSPQEQLP---FIINFTLSPKSALPEPOLYFPA--FGQNDKTIAGELATPFESRGWG 371
DB	66 DDETLKPSPLPCAGVIFNFEIWPFGADKPNPKIYLPICAVYTKGDKLDIADGMDSPFKQGS 125
QY	372 GLAKSPAD-LASYYPDVLQTAHNLQAWISFYSYKGGKPKYMSVY 414
DB	126 KSFHSYKNDYIKAFVXGKVMCRHH---DISFSYKGGAYITAY 166
RESULT 6	
ID	Q8S922 PRELIMINARY; PRT; 577 AA.
AC	Q8S922;
DT	01-JUN-2002 (TrEMBLrel. 21, Created)
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Cell wall invertase (EC 3.2.1.26).
GN	OSCINI.
OS	Oryza sativa (japonica cultivar-group).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC	Ehrhartoideae; Oryzaceae; Oryza.
OX	NCBI_TaxID=39947;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ev. Nipponbare; TISSUE=Panicle;
RX	MEDLINE=21975747; PubMed=11978873;
RA	Hirose T., Takano M., Terao T.;
RT	"Cell wall invertase in developing rice caryopsis: Molecular cloning
RT	of OSCINI and analysis of its expression in relation to its role in
RT	grain filling.";
RL	Plant Cell Physiol. 43:452-459 (2002).
CC	-!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR	EMBL; AB073749; BAB90855.1; --
DR	Gnamene; Q8S922; --
DR	GO; GO:0004564; F:beta-fructofuranosidase activity; IEA.
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR	InterPro; IPR001362; Glyco_hydro_32.
DR	Pfam; PF00251; Glyco_hydro_32; 1.
DR	SMART; SM00640; Glyco_32; 1.
DR	PROSITE; PS00609; GLYCOSYL HYDROL_F32; 1.
KW	Glycoprotein; Glycosidase; Hydrolase.
SQ	SEQUENCE 577 AA; 63984 MW; 41FE6384284F97F1 CRC64;
Query Match 5.0%; Score 117; DB 10; Length 577;	
Best Local Similarity 24.7%; Pred. No. 0.41;	
Matches 111; Conservative 50; Mismatches 152; Indels 136; Gaps 29;	
QY	62 REVIIIPALGPY--PEKGQPM-----HWKSHL-----TRGLPFELSFNYSK 100
DB	165 REMVKPAXNPVATPEPGNNATQFRDPTTAVADGHRMLVGLKGLAGLAYLYRSRDFK 224
QY	101 SLLRPAPFPLGS-LTGTKDDP---FNTQAIRVLQ-DLKAWVP-----GLDLWFDFH 148
DB	225 TWVR-AKHLHSALTGMWCEPDPFPLQA--FGLQGLDTSVPSSKYVLKNSLDLTRYDY 281
QY	149 TKALVUSSEEARLLDRDIEIPVFTONKLAADLEPSGD-----IVLKYIYYP- 196
DB	282 T-----VGIYNKVTYRVPDNPAGDYHRLRYDYGNFYASKTFFDFV 322
QY	197 RIKSTATGTPKRLMFDALKAADKFGKVATPLAI-----LEEFIAERAPTLLGHFLS 248
DB	323 KHRILLGWANESDSVTYDKAKGNAGIHAIPKRWLDPSGKQLLOWFEELETLRGKSVS 382
QY	249 C--DLVKPSE---SRIKYV--CMERQDLASIEGIWTLNGRRNDPDTLGLDAL----- 295
DB	383 VFDKVVKPGHEHFOVTLGTYQADVEVLSVGLXKAEAL-----DPAFGDNAERLCGAKG 437

QY 296 -----RELWQLPVTEGICPLPNCF---YEGTSPQEQLPPIINFITLSPKSALPEPQ 344
 DB 438 ADVRGVVGGLWL--ASAGLEEXTAVFRVFKPAGHAE--PVVLMCTDPTKSSL-SPD 492
 QY 345 IYFPAP-GONDKTIAEGLATF-----PESRWGG-----LAKSYPA-----DLASY-- 385
 DB 493 LYKTFAGVTDIDISSKISLRSIDRSVVSFGAGGKTCILSRVPSMAIGDKAHLYVF 552
 QY 386 --PDVDTQANHLQAWISFSYKGGKPYMS 412
 DB 553 NNGEADIK-ISHLKAW-----EMKKPLMN 575
 RESULT 7
 Q96X17 PRELIMINARY; PRT; 1772 AA.
 AC Q96X17;
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Sec7p.
 GN SEC7.
 OS Pichia pastoris (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4922;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Soderholm J.F., Bevis B.J., Glick B.S.;
 RT "A vector for pop-in/pop-out gene replacement in Pichia pastoris.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR ENBL: AF356651; AAK40234.1;
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR00904; Sec7.
 DR Pfam; PF01369; Sec7; 1.
 DR SMART; SM00222; Sec7; 1.
 DR PROSITE; PS50190; Sec7; 1.
 SQ SEQUENCE 1772 AA; 200298 MW; 3ED09DF4136C8888 CRC64;

Query Match 4.8%; Score 112; DB 3; Length 1772;
 Best Local Similarity 20.4%; Pred. No. 5.8;
 Matches 95; Conservative 62; Mismatches 150; Indels 159; Gaps 23;
 QY 49 YDHAQYKFLCIHREVIIPALGPYKPGOPMHWKSHLTFGLP-FELS----FNYSKSL 103
 DB 525 YDCDSQLPSLC---EGLI-----DYLTFSTLTVSEISQOQKINFRASLT 565
 QY 104 R-----FAPEPLGSLTGKDDPN-----TQAIRPVQLDKAMVPL 140
 DB 566 RSLAVYSLKQSPMLSKLGGANVPDPEASYNPFGEYAVIESIECVLVLSLST----- 620
 QY 141 DLEWDFHTK-ALVWSEEEARTLDRIDIPVFKTONKLAADL-BPSGDIVLKTVIYPRI 198
 DB 621 ---WVDSVAKQAVESEEDTALSNGAGEDEILSORSESATOLSETG-----IPDP 670
 QY 199 KSIATGTPKRLMFDIAKAADKFGKVATPLAILEEPIAERAP-----TLL 243
 DB 671 AKFDTOKRKATLFCQAFNYKPKIGIAKAIKESGFIKDDSPQETAKELLVTDGLDKTI 730
 QY 244 GHFLSCDLVKPSRSIKVYCMERQDLASIEGIWTLNGRNDPETHLDGLDALRELWQ--- 300
 DB 731 GEYLG-----EGDEKNITI---MDFVLDLMDFSG-----LEFVEAMRTFLQNR 771
 QY 301 -----LLPVTEGLCPLPNCFYEPGTSPQQLPPIINFITL-----SPKSALEPQ 344
 DB 772 LPGESQKIDRFMLKFAERF-----VLNNGPTFANADVFLVLAISVILLNTDQHSQVKR 826
 QY 345 IYFPAGGONDKTIAEGLATFFESRWGGGLAKSYPA--YYP-----DVDLOT----- 392
 DB 827 MSLOQFIRNAGIDG-----QDLPETLSKIYFYIQSNEIKLOEQQAAL 872
 QY 393 -ANHLQ-----AWISFSYKGGKPYM-----SVYLHTFEAFSAAG 427

DB 873 LAGHIQPEPVTGLFSFRNQEREQYMLSKELTLNTEKVKFSFGQE 917
 RESULT 8
 Q8TXF1 PRELIMINARY; PRT; 462 AA.
 AC Q8TXF1;
 DT 01-JUN-2002 (TRENBLrel. 21, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE CO dehydrogenase/acyl-CoA synthase gamma subunit (Corrinoid Fe-S protein).
 DE protein.
 GN CDHE OR MK0723.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 OC Methanopyrus.
 OX NCBI_TaxID=2320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozayavkin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
 DR EMBL; AE010365; AAM01937.1; -;
 DR GO; GO:0006730; P-one-carbon compound metabolism; IEA.
 DR InterPro; IPR004486; CdhD.
 DR InterPro; IPR007202; Fes.
 DR Pfam; PF03599; CdhD; 1.
 DR Pfam; PF04060; Fes; 1.
 KW Complete proteome.
 SQ SEQUENCE 462 AA; 50852 MW; 1E97B54407368D3B CRC64;

Query Match 4.8%; Score 110.5; DB 17; Length 462;
 Best Local Similarity 27.0%; Pred. No. 1;
 Matches 62; Conservative 28; Mismatches 91; Indels 49; Gaps 12;
 QY 125 AIRVLQDLKAMVPLGLEWDFHTKALVWSEEEARTL--LDRDIEIPVFKTONKLAADL 182
 DB 159 AVALITTDPKVMEAGLDV--FDERPLYPATEENVEDLAKLAADGCPPL-----GLHARDV 212
 QY 183 EPSGDIVLKTVIYPRISIKSIATGPKERLMFDAIKAADKFGKVATPLAILEEPIAERAPTL 242
 DB 213 EDLVPLVVEAQYVTDLLDLPGT--BFGPHDVVSTTDKLAETIRK--AAIEEFESFGYPTL 268
 QY 243 LG---HFLSCDLVKPS--ESRIKVKMERQDLASIEGIWTLNGRNDPETHLDGLDALR 296
 DB 269 VTFPFAFLDEDDPVKAARRESYLASACVRYADILIMDV----- 308
 QY 297 ELMQLLPV-TEGLC-----FLPNCFYEPGT-----SPOQLPFI--NFTLS 335
 DB 309 EPWALLPVLTRQCVVTDPREPQVEPGLYRIGDPDNPSPVLVTTNFTLT 358

RESULT 9
 Q21549 PRELIMINARY; PRT; 490 AA.
 AC Q21549;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE M176.2 protein.
 GN M176.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.


```

DR GO; GO:0005351; F:sugar porter activity; IEA.
DR GO; GO:0009401; P:phenololipyrivate-dependent sugar phospho. . .; IEA.
DR InterPro; IPR002114; HPr Serp S.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007091; LRR_Rninh.
DR InterPro; IPR003590; LRR_Rninh_sub.
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR004020; PAAD_DAPIN_dom.
DR Pfam; PF00560; LRR_2.
DR Pfam; PF02758; PAAD_DAPIN; 1.
DR SMART; SM00368; LRR_R1; 9.
DR PROSITE; PS50824; DAPIN; 1.
DR PROSITE; PS50837; NACHT; 1.
DR PROSITE; PS00589; PTS_HPR_SER; 1.
SQ SEQUENCE 1043 AA; 118910 MW; C321FBBCC65206DFF CRC64;

Query Match 4.4%; Score 102; DB 4; Length 1043;
Best Local Similarity 22.3%; Pred. No. 19;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;

QY 79 MHWKSHL---TRGLPELSFNYSKSLRPAPEPLGSLGTRK-DPPFNTQAIRPVLDLK 134
DB 250 LHWANGVLFQORFYSYFYLSCHIRYKETTFAELISLDWPDFDAPIEFMSQP--EKLL 307

QY 135 AMVPEGLDLEWDFHTKALVSEBEARTLLDRD-----IEIPVK-TQNKLAADLEPSGD 187
DB 308 FIIDGFE-----EIIISERSBSLDDGSPCTDWYQELPVTXILHSLKKELVPLAT 358

QY 188 --IVLXTIYPRKSI-----ATGPKERLMFDAIYAADKFGKVAIPLAIEE---- 233
DB 359 LLTIKTITFWRDUKASLVNCPVQITGFTGDDLRVVFMEHFDSDSEVEKILQLRKNETL 418

QY 234 FIARAAPTLLGHFLSC-----DLVKPSSRIKVC-----MERQLDLA----- 271
DB 419 FHSCSAPMVCWTVCCLQKQKVRYDLSQITQTTSLYAYFNSLFFSTAEVDLADDSWFG 478

QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
DB 479 QWRALCSLAIEGLWSNFTFNKEDTIEGLEVPFIDSLYEFNLQKINDCGGCTFTHLS 538

QY 312 -----PNCFEYEGTSPOQLPFIINFTLSPKSALPPEQIYFPAGQNDKTI 357
DB 539 FQEFFPAMSEVLEPEPFPHSTKPOE-MKNLLQVLLDKEAVTVPVLF-FFGLLNKNI 596

QY 358 AEGL 361
DB 597 AREL 600

RESULT 12
Q8K0C4 PRELIMINARY; PRT; 503 AA.
AC Q8K0C4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome P450, 51.
GN CYP51.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary Gland;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC TISSUE=Extraembryonic tissue, Placenta, Skin, and Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.

```

```

RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; BC031813; AAH31813.1; -
DR EMBL; AK028355; BAC25900.1; -
DR EMBL; AK028815; BAC26134.1; -
DR EMBL; AK076983; BAC36548.1; -
DR MGD; MGI:106040; Cyp51.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PRO0385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 503 AA; 56775 MW; 6A4BBA350F1D85C7 CRC64;

Query Match 4.3%; Score 101; DB 11; Length 503;
Best Local Similarity 19.1%; Pred. No. 7.6;
Matches 97; Conservative 68; Mismatches 196; Indels 148; Gaps 21;

QY 10 LLPFPFVVLQALNLSNKDHTKMYSTAPMFATMAGAGYDVHAQYKF-LCIHREVIIP- 67
DB 6 LQGGVWLQAME-----QVTGNNLSLTLLIACFTLSLVYLFRLAVGHVQLPA 56

QY 68 -ALGPYEPKQPMHWKSHLTRFG-LPELSFN-YSKSLRFAFEPLGS----- 112
DB 57 GAKSP-PHIYSPFPLGHAFAGKSPLEENAYEKVPVFTWVGKTFYLLGSDAAA 115

QY 113 -LTGTDKDDPNTQAI-----RPV-----LQDLKAMVPGLDLEWDFHTKAL 152
DB 116 LFNKSKVDLNAEEVYGRLTTPVFGKVAYDVNPAIFLEQKKIKSLNTA---HFKQYV 172

QY 153 VVSEEAR-----TLDRDIEIPVFTQNKLAADLEP 184
DB 173 PIIEKEAKEYFQSWGESGERNVFEALSELILLTASHCLHGKEIRSQLEKVAQYADLD- 231

QY 185 SGDILVLTYYIP-----RIKSI-----ATGPKERLMFDAIKAAD 219
DB 232 -GGFTHAALLPAWLPJPSFRRDRRAHREIKNIFYKAIQKRRLSKPEAFEDILTLLDSTY 290

QY 220 KFGKVATP-----LAILEEFAIERAPTLLGHFLSCDLVKPSSRI---KVYCM- 265
DB 291 KGRPLTDEISGMLGLLAGOHTSTTSWNGFFLAKD--KPLQEKVLEQKAVCGED 348

QY 266 -----RQLDASIE-----GIWTLNGRNDPETLDGLDALRELWQLLPVTEG 307
DB 349 LPPLTYDQLKDLNLDRCIKETLRLRPIMTMWMAKTPTQVAG-----YTIPIPGHQ 400

QY 308 LCPLPNCFYEGTSPQEQLPFIINFTLSPKSALPPEQIYFPAGQNDKTIAGLTFPES 367
DB 401 VCVSPTVQRLKDSWAERLDFNPDRLQDNPAEGEKFAYVFPFAGRHRGCVGENFAIVQIK 460

QY 368 RGMGGLAKSYFADLAS-YYPDVLDTQANH 395
DB 461 TINSTMLRYEFLDINGYFPYVNTYTMIH 489

RESULT 13
Q8BSQ7 PRELIMINARY; PRT; 503 AA.
AC Q8BSQ7
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytochrome P450.
GN CYP51.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
ON [1]_RN
OX RN
XP SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Forelimb;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
"Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RT Nature 420:563-573 (2002).
RL EMBL; AK031059; BAC27231.1; -.
DR MGD; MGI:106040; Cyp51.
DR GO; GO:0004295; F-tyrosin activity; IEA.
DR GO; GO:0006118; P-electron transport; IEA.
DR GO; GO:0006508; P-proteolysis and peptidolysis; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PS00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR SEQUENCE 503 AA; 56745 MW; 651B4F350B4DC0C7 CRC64;
SQ

Query Match 4.3%; Score 101; DB 11; Length 503;
Best Local Similarity 19.1%; Pred. No. 7.6;
Matches 97; Conservative 68; Mismatches 196; Indels 148; Gaps 21;

QY 10 LLPKPYVLVSQALNSLNKDKTKWYSTAPMFATMAGAGYDVHAQYKF-LCIHREVIIP- 67
Db 6 LLQSGGVWLQAME-----QVTGNNLSTLLIACAPTLSLVYLFRLAUGHVWLPA 56
QY 68 -ALGPPYKQOPMHWKSHLTRFG-LPELSFN-YKSLRLFAPEPLGS----- 112
Db 57 GAKSP-PHYSPYPIFFGHAIAPGKSPFIEFNAYEKYGPVFTMWGKTFYLLGSDAA 115
QY 113 -LTGTKDDPNTQAI-----RPV-----LQDLKAMVPGLDLEWDFHFTKAL 152
Db 116 LFNSENEDLNAEVYGRLLTFYFGKVAYDVPNAIFLEQKKIIXKSLNIA---HFQYV 172
QY 153 VVSEEAR-----TLDRDIEIPVFKTONKLAADLEP 184
Db 173 PIIEKAKEYFQSGWSEGERNVFEALSELIIITASHCLHGKEIRSQ-NERVAQIYADLD- 231
QY 185 SGDIVLKTYYP-----RIKSI-----ATGTPKRLMFDKAAAD 219
Db 232 -GGFAHAAILLPAWLPLPSFRDRRAHREIKNIFYKAIQKRLSKEPAEDIIQLTLDSTY 290
QY 220 KFGKVATP-----LAILEFIAERAPYTLIGHFLSCDLVLPSESRI----KVYQWE- 265
Db 291 KDCRPLTDEISGMILIGLLAGQSTSTTSAAWGFFLAKD--KPLQKCYLEQKAVCGED 348
QY 266 -----RQDLASIE-----GIWTLNGRNDPETLQGLDALRELWQLLPVTEG 307
Db 349 LPPLTYDQLKDLNLLORCIKETURLRPPIIMTMRMAKTPQTVAG-----YTIPPGHQ 400
QY 308 LCLPNCFYEPFGTSPQOLFPIINFITLSPKSAALPEQIYFPAPQNDKKTIAEGLATPFES 367
Db 401 VCVSPVTNQRKDSWAERLDFNPDRYLQDNPASEKEFAYVFPFGAGRHCRCVGFAYVQIK 460
QY 368 RGHGGLAKSYPADLAS-YYPDDVLDQTANH 395
Db 461 TIWSTMLRLYEFDLNGYFFFTVNTYTMIH 489

RESULT 14
Q42710
ID Q42710 PRELIMINARY; PRT; 878 AA.
AC Q42710;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

```

DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DR	Lipoxxygenase [EC 1.13.11.12].
DS	Cucumis sativus (Cucumber).
QC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
CC	eurosid1; Cucurbitales; Cucurbitaceae; Cucumis.
CC	NCBI_Taxid=3659;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Cotyledon;
RA	MODLINE=97054584; PubMed=8898881;
RX	Hoehne M., Nellen A., Schwennesen A., Kindl H.;
RT	"Lipid body lipoxxygenase characterized by protein fragmentation, cDNA
RT	sequence and by its very early expression during germination of
RT	cucumber seeds.";
RL	Eur. J. Biochem. 241:6-11 (1996).
DR	EMBL; X92890; CAA63483.1; -.
DR	PIR; S74207; S74207.
DR	HSP; P08170; 2SBL.
DR	GO; GO:0005506; F:iron ion binding; IEA.
DR	GO; GO:0016165; F:lipoxxygenase activity; IEA.
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.
DR	GO; GO:0006118; P:electron transport; IEA.
DR	InterPro; IPR000907; Lipoxxygenase.
DR	InterPro; IPR001024; Lipoxxygenase_LH2.
DR	InterPro; IPR008976; PLAT LH2.
DR	Pfam; PF00305; lipoxxygenase_1.
DR	Pfam; PF01477; PLAT; 1.
DR	PRINTS; PR00087; LIPOXYGENASE.
DR	SMART; SM00308; LH2; 1.
DR	PROSITE; PS00711; LIPOXYGENASE_1; 1.
DR	PROSITE; PS00081; LIPOXYGENASE_2; 1.
DR	PROSITE; PS50095; PLAT; 1.
DR	Oxidoreductase.
KN	SEQUENCE 878 AA; 99768 MW; A92660AF4D1ED3BE CRC64;
SQ	
	Query Match 4.3%; Score 101; DB 10; Length 878;
	Best Local Similarity 21.7%; Pred. No.17;
	Matches 94; Conservative 49; Mismatches 129; Indels 162; Gaps 26;
Qy	61 HREVIIPALG-----PYPEK---GQPMHWKSHLTREGLPFELSFNYKSLLRPAPEPLGS 112
Db	235 HR-----PILGGTTEYPYPRGRTGRPSRRDH-----NYESRL-----SPINS 273
Qy	113 LT--GTKDDPFN-----TQAIRFVLQDKAMVPGLDLEWDFHFTKALVVS 155
Db	274 LDIVYKDFNGFHLKMSDFLGYTLKALSIISIKFGLSIFDVTP-----NEFDNF----- 322
Qy	156 EEPARTLLDRDIEIP--VPKTONKLAADLEPSGDIVLKYIYIPRIKSIATGTPKRLMFD 213
Db	323 -KEVDNLFRRGFPPIPNFAKT---ITEDLTP-----PLFKALVRNDGKFLKFP 367
Qy	214 AIKAAKFGKV--ATELAILEBFIARPTLGHFLSCDLVXPSERIIVYQWEROGLDLA 271
Db	368 TPEVV-KDNKIGWSTDEEFAREMLAGPNLLI-----RRLEAFPTSKLDP- 412
Qy	272 SIEGIWTLNRRNDPPT-----LDGLDALRELWQ-----LIPVTEGLCLPLNCFYEPGT 320
Db	413 -----NVYGNQNSTITEESHKIGLDGLTVDEAMQNRLYIVDFHDALMPYLTRMNATST 466
Qy	321 SPOEQLPFIH---NFTLSP---KSALPERQ-----IYEPAGQNDKTTAEGLATFF 365
Db	467 KTVATETLLLLKDDGTLKPLVIELALPHPGQDOLGAISKLYFPAENGVOKSI----- 518
Qy	366 ESRGWGLAKSPADLASVYPVDQLQTANHQAQWISFSYKGGPKYMSVYLHT---FEAFS 422
Db	519 ----W-QLAKEY-----VTVNDGVYHQLISHLWHTHAVLEFPV 551
Qy	423 AAA-QEVAMCHDH 435
Db	552 IATHRLQSLVLPHT 565

